PRODUCTION AND CHARACTERIZATION OF RECOMBINANT FIBRILLOGENIC PROTEINS AND POTENTIAL APPLICATIONS OF PROTEIN SELF-ASSEMBLY IN NANOBIOTECHNOLOGY

Fulvio Guglielmi

Dottorato in Scienze Biotecnologiche – XX ciclo
Indirizzo Biotecnologie Molecolari
Università di Napoli Federico II

Dottorato in Scienze Biotecnologiche – XX ciclo
Indirizzo Biotecnologie Molecolari
Università di Napoli Federico II

PRODUCTION AND CHARACTERIZATION OF RECOMBINANT FIBRILLOGENIC PROTEINS AND POTENTIAL APPLICATIONS OF PROTEIN SELF-ASSEMBLY IN NANOBIOTECHNOLOGY

Fulvio Guglielmi

Dottorando:	Fulvio Guglielmi
Relatore:	Prof.ssa Renata Piccoli
Coordinatore:	Prof. Giovanni Sannia

Ai miei genitori,
a Emilia

INDEX

ABBREVIATION

ApoA-I	Apolipoprotein A-I
[1-93]ApoA-I	The fibrillogenic N-terminal 93-residue polypeptide of ApoA-I
GST	The gluthatione S-transferase from *S. japonicum*
GST-[1-93]ApoA-I	The chimeric protein containing [1-93]ApoA-I fused to GST
GFP	Green Fluorescent Protein
mnei	Recombinant single chain monellin
oryc	Cystatin I from *Oryza sativa L. japonica*
BSA	Bovine serum albumin
DTT	Dithiotreitole
RP-HPLC	Reverse Phase High Performance Liquid Chromatography
IEX-C	Ion Exchange Chromatography
LB medium	Luria-Bertani medium
O.D.	Optical density
PCR	Polymerase Chain Reaction
PVDF	Polyvinylidene fluoride
SDS	Sodium-dodecyl-sulfate
SDS-PAGE	Gel-acrylamide electrophoresys in denaturing conditions
Tris	Tris-hydroximetyl-amino methane
AFM	Atomic Force Microscopy
EM	Electron Microscopy
ANS	8-anilino-1-naphthalenesulphonate
ThT	Thioflavin T
CDNB	1-chloro-2,4-dinitrobenzene
CR	Congo Red
RCT	Reverse Cholestérol transport
TFA	Trifluoroacetic acid

SUMMARY

An increasing number of human diseases is linked to protein misfolding and aggregation in amyloid fibrils. Among amyloidogenic proteins, 13 mutated versions of apolipoprotein A-I (ApoA-I) have been associated to amyloid diseases. Fibrils are mainly constituted by N-terminal fragments of ApoA-I, about 90-100 residue long. We recently set up a reliable expression system to produce a recombinant version of 1-93 polypeptide (named [1-93]ApoA-I) as a pure and stable product. Conformational analyses of [1-93]ApoA-I revealed that under acidic conditions the polypeptide rapidly converts from random coil to an unstable intermediate with a helical/molten globule state, able to bind ANS, as described for the ex vivo isolated polypeptide. This species is thought to be a key early intermediate in the fibrillogenic pathway, leading to the formation of a β-sheet-based polymeric structure, having amyloid-like features, i.e. the ability to bind specific dyes, such as Congo Red (CR) and Thioflavin T (ThT). AFM analyses confirmed that [1–93]ApoA-I is able to generate typical amyloid fibrils.
Furthermore, we were also able to express the 8 amyloidogenic variants of [1–93]ApoA-I identified so far in patients affected by amyloidosis. The availability of the complete panel of pathological ApoA-I polypeptides opens the way to a detailed study, presently in progress, aimed at the definition of the impact of each mutation on protein aggregation propensity. Preliminary data indicated that almost all mutants display an increased aggregation propensity with respect to [1-93]ApoA-I.
During my stage at the Medical Research Council (MRC) in London, I extended my studies to plant cystatins, i.e. protease inhibitors, largely homologous to human cystatins, some of which are involved in amylopathies. Hence, plant cystatins represent an interesting model system to shed light on cystatins aggregation. Two cystatins were analysed: mnei, a recombinant engineered monellin, and oryzacystatin I (oryc), a cystatin from *O. sativa*. I found that, under acidic conditions, both proteins undergo a temperature-driven transition from β–sheet to random coil, that eventually leads to aggregation. Mnei was found to generate typical amyloid fibrils, able to bind the amyloid specific dye ThT. No amyloid-like structures were instead found to be present in oryc aggregates, whereas stable trimeric species were isolated and supposed to prevent protein self-aggregation.
I also planned to engineer fibrillogenic polypeptides to produce new bio-materials obtained by molecular self-assembly. Amyloid fibrils are in fact good candidates to construct nanostructured materials, by means of their ordered, polymeric structure. Catalytic fibrils were then designed, using a fusion protein, made up of [1-93]ApoA-I linked to GST enzyme, as the monomeric unit in a fibrillogenic process. The former will act as the fibrillogenic moiety, the latter as the catalytic moiety. The fusion protein was expressed and isolated as a pure product, and found to retain both the native fold and the catalytic activity of GST enzyme. Moreover, upon incubation under slightly acidic conditions, the fusion protein was found to bind ANS, ThT and CR dyes, indicating that a pH-driven transition, followed by aggregation in amyloid-like structures, occurred. When analysed by AFM, typical protofilaments were detected. The high molecular weight fibrillar material, immobilised on filter, was then tested for enzymatic activity, by providing in continuous the reaction mixture containing the GST substrates. Fibrils were found to be catalytically active.

RIASSUNTO

Lo scopo del mio lavoro sperimentale ha riguardato lo studio delle proteine fibrillogeniche, una classe di proteine propense ad aggregare in strutture ordinate, le fibrille amiloidi, associate a gravi patologie definite amiloidosi. Per dar luogo al processo fibrillogenico, tali proteine hanno accesso a conformazioni alternative parzialmente destrutturate; per tale motivo le patologie amiloidi sono anche definite malattie confomazionali, o da *misfolding* proteico.
Negli ultimi anni nuovo interesse è sorto intorno al tema del *folding* proteico. Numerose evidenze, tra cui l'esistenza di proteine *natively unfolded*, cioè proteine parzialmente destrutturate anche in condizioni fisiologiche, hanno suggerito un panorama più complesso nell'ambito dello studio del *folding* proteico. Rispetto alla visione classica secondo la quale ad una struttura corrisponde una funzione, una interpretazione più ampia del processo di *folding* proteico prevede per ogni proteina l'esistenza di un profilo energetico a più stati, popolati da intermedi transienti solo parzialmente strutturati. La proteina esplora tale profilo energetico per raggiungere la propria conformazione nativa. La "complessità" del profilo energetico può rendere ragione sia del numero diverso di conformazioni che una proteina può assumere, sia della velocità con cui essa evolve verso lo stato nativo. Tale modello prevede la possibilità che una proteina generi stabilmente oligomeri o aggregati di diverso tipo, quando alle interazioni intramolecolari si sostituiscono quelle intermolecolari. Gli aggregati proteici si configurano quindi come prodotti del processo di *folding* proteico.
Eventi destabilizzanti, quali le condizioni del microambiente o la presenza di mutazioni, possono essere responsabili della formazione di intermedi proteici parzialmente destrutturati. Tali intermedi espongono, o giustappongono, superfici nascoste nella struttura nativa, o distanti nello spazio, permettendo l'aggregazione di monomeri in strutture polimeriche.
Le fibrille amiloidi costituiscono aggregati proteici ordinati caratterizzati da una struttura polimerica detta β-*cross*, in cui le molecole della proteina, o sue regioni fibrillogeniche, assumono una conformazione a foglietti β disposti perpendicolarmente all'asse della fibrilla. L' osservazione che il medesimo motivo β fibrillare sia assunto da proteine non correlate tra loro per struttura o funzione, e che numerose proteine possono generare fibrille *in vitro*, suggerisce che la struttura fibrillare sia una forma di organizzazione superstrutturale preservata durante l'evoluzione. I depositi proteici generati da proteine distinte, condividendo specifiche caratteristiche strutturali, mostrano proprietà comuni, quali la capacità di legare specifici coloranti, quali la Tioflavina T (ThT) o il Congo Rosso (CR).
Un numero sempre crescente di patologie nell'uomo è correlato alla elevata propensione di alcune specifiche proteine ad aggregare generando depositi amiloidi in organi o tessuti. Tra queste patologie annoveriamo il morbo di Parkinson, il morbo di Alzheimer e le encefalopatie spongiformi.
Più di 20 proteine nell'uomo sono state associate ad amiloidosi: tra queste la Apolipoproteina A-I (ApoA-I), il principale costituente della frazione delle lipoproteine ad alta densità del siero (HDL). La presenza di specifiche mutazioni nel gene ApoA-I genera varianti della proteina riscontrate in pazienti affetti da patologie amiloidi. Tredici varianti sono state finora descritte. Nei pazienti affetti da tali amiloidosi, le fibrille sono costituite dalla porzione N-terminale di ApoA-I, di lunghezza variabile

compresa tra 80 e 100 amminoacidi. Tali polipeptidi sono generati attraverso un meccanismo ancora sconosciuto.
Delle tredici mutazioni associate a patologie, 8 sono localizzate nella regione N-terminale della proteina corrispondente al frammento fibrillogenico, e pertanto definite mutazioni "interne"; 5 mutazioni invece sono localizzate a valle di tale regione (mutazioni "esterne"). Studi effettuati dal gruppo del Prof. V. Bellotti dell'Università di Pavia sul polipeptide costituito dai residui 1-93 di ApoA-I, estratto da fibrille isolate dal cuore di un paziente sottoposto a trapianto, hanno evidenziato che tale polipeptide si può definire *natively unfolded* in quanto destrutturato in condizioni fisiologiche.
La impossibilità di ottenere sufficiente materiale *ex vivo,* nonché le difficoltà legate alla produzione in forma ricombinante di un polipeptide destrutturato, hanno finora limitato la possibilità di analizzare in dettaglio il processo di fibrillogenesi di ApoA-I.
Nell'ambito del mio progetto di ricerca, ho contribuito a mettere a punto un efficace sistema di espressione, che permette di ottenere il polipeptide ricombinante in forma pura e stabile, in quantità sufficienti per studi strutturali e funzionali.
Il cDNA codificante la regione N-terminale di ApoA-I (residui 1-93), clonato in un opportuno vettore, è stato espresso come prodotto di fusione con una proteina batterica, altamente stabile e solubile, quale l'enzima glutatione S-transferasi (GST da *S. Japonicum*). La proteina di fusione, isolata mediante cromatografia di affinità, è stata poi sottoposta a proteolisi sito-diretta per generare il polipeptide di interesse. Quest'ultimo è stato poi isolato all'omogeneità mediante cromatografia a fase inversa. Il prodotto ricombinante è stato denominato [1-93]ApoA-I.
L'analisi conformazionale del polipeptide [1-93]ApoA-I, effettuata mediante dicroismo circolare nel lontano ultravioletto, ha dimostrato che in condizioni fisiologiche [1-93]ApoA-I è una proteina *natively unfolded*. Abbiamo inoltre dimostrato che, a seguito della acidificazione della soluzione (pH 4), il polipeptide subisce una transizione conformazionale da una struttura *random coil* ad una struttura ad α-elica, come riscontrato anche per il polipeptide estratto *ex vivo*. Tale transizione è risultata reversibile se le condizioni fisiologiche vengono ripristinate.
Esperimenti di spettrofluorimetria, effettuati utilizzando il fluoroforo 8-anilino-1-naftalene-solfonato (ANS), hanno indicato che tale composto lega in maniera reversibile il polipeptide a pH 4, ciò che indica la presenza di una struttura di tipo *molten globule*. Sulla base di questi dati abbiamo ipotizzato che la struttura α-elicodale, con caratteristiche di *molten globule,* è un intermedio chiave nel processo di fibrillogenesi e che la sua formazione rappresenta un evento precoce e reversibile in questo processo.
Abbiamo poi dimostrato che la prolungata esposizione del polipeptide in ambiente acido induce la conversione dell'intermedio α-elicodale in una struttura prevalentemente del tipo β-sheet. La formazione di tale struttura viene seguita da aggregazione proteica, come dimostra la scomparsa del segnale di dicroismo circolare nel tempo. La caratterizzazione degli aggregati formati dal polipeptide in condizioni acide, attraverso l'utilizzo di coloranti specifici per fibrille amiloidi, quali Tioflavina T e Congo Rosso, ha evidenziato la presenza della caratteristica struttura amilode *β-cross*. In tali analisi l'abbassamento del valore di pH emerge come un evento destabilizzante, in grado di favorire cineticamente la formazione delle fibrille.
Dai dati riportati è risultato chiaro che le proprietà riscontrate per il polipeptide ricombinante sono sovrapponibili a quelle della sua controparte nativa, ciò che avvalora l'utilizzo del polipeptide ricombinante come modello per lo studio del processo di fibrillogenesi.

Ad avvalorare quanto riportato, analisi *in silico*, effettuate con i comuni algoritmi predittivi, hanno mostrato un elevato potenziale fibrillogenico del polipeptide, tanto più elevato quanto più acido è il pH. Da tali analisi, la regione compresa tra i residui 13-25 del polipeptide è risultata presentare il più elevato potenziale fibrillogenico.
Esperimenti di gel-filtrazione hanno inoltre mostrato che il polipeptide in condizioni fisiologiche assume una struttura dimerica. Tali dati sono in accordo con risultati precedentemente ottenuti mediante esperimenti di proteolisi limitata, accoppiata a spettrometria di massa, condotte dal gruppo del Prof. P. Pucci dell'Università di Napoli.
Infine, dati di microscopia a forza atomica (AFM) hanno dimostrato che il polipeptide ricombinante, a seguito dell'incubazione in condizioni di pH acido, è in grado di generare tipiche fibrille amiloidi, morfologicamente simili alle fibrille formate dal polipeptide naturale, e prodotte in una scala temporale paragonabile.
L'insieme dei dati ottenuti indica l'esistenza di un complesso processo fibrillogenico a più stadi, attraverso cui il polipeptide [1-93]ApoA-I assume diverse conformazioni e genera fibrille amiloidi. In tale processo sembra essere determinante il delicato equilibrio esistente tra le interazioni stabilizzanti la forma nativa e le interazioni in grado di favorire l'aggregazione del polipeptide.

I risultati ottenuti nella caratterizzazione del processo fibrillogenico del polipeptide [1-93]ApoA-I ci hanno incoraggiato a produrre in forma ricombinante anche le varianti del polipeptide riscontrate nei pazienti affetti da amiloidosi.
Pertanto, ho ottenuto il cDNA codificante ciascuna delle 8 varianti del polipeptide fibrillogenico, attraverso esperimenti di mutagenesi sito-diretta del cDNA codificante [1-93]ApoA-I. Tra queste varianti, 6 contengono una mutazione puntiforme (G26R, T50R, L60R, L64P, L75P e L90P), mentre 2 contengono una delezione della sequenza 70-72 e 60-71, rispettivamente (nell'ultimo caso è presente anche l'inserzione dei residui VT). Ciascun cDNA è stato clonato vettore di espressione già utilizzato per il polipeptide [1-93]ApoA-I, a valle della sequenza codificante l'enzima glutatione-S-transferasi. Seguendo il protocollo precedentemente riportato, ho espresso e isolato all'omogeneità le 8 varianti patologiche, ottenendo in questo modo il set completo dei mutanti del polipeptide.
L'ottenimento delle varianti del polipeptide amiloidogenico in forma ricombinante risulta essere l'unica strategia utilizzabile per poter affrontare studi strutturali e funzionali sulle varianti stesse. Sono attulmente in corso analisi sulla dinamica conformazionale e sul potenziale fibrillogenico delle varianti del polipeptide [1-93]ApoA-I. Dati preliminari hanno dimostrato che le varianti hanno una elevata tendenza a generare fibrille, anche più elevata di quella del polipeptide non mutato. Tali analisi sono volte a definire l'impatto delle singole mutazioni sul potenziale fibrillogenico del polipeptide, e sono un utile strumento per l'identificazione dei determinanti molecolari della fibrillogenesi.

Nell'ambito dello studio sulle proteine fibrillogeniche, è parso di interesse effettuare studi su alcune cistatine da piante, come proteine modello per lo studio del processo di fibrillogenesi delle cistatine umane associate ad amiloidosi. Infatti le cistatine, inibitori delle proteasi a cisteina, rappresentano un altro esempio di proteine correlate a severe amilopatie. Tra esse annoveriamo la cistatina C, responsabile della angiopatia amiloide ereditaria (HCCAA) e la cistatina B implicata nella malattia di Unverricht- Lundborg. In entrambi i casi, mutazioni nella sequenza proteica sono identificate come il fattore determinante la patologia.

Tali studi sono stati effettuati durante il mio soggiorno al Medical Research Council (MRC) di Londra, presso il laboratorio della Dott. A. Pastore.
Le cistatine da pianta (fitocistatine) possono, in particolari condizioni, generare *in vitro* fibrille amiloidi. Inoltre esse condividono con le cistatine umane il tipico motivo strutturale delle cistatine, consistente in una singola α-elica intorno alla quale cui si dispone un foglietto β a 5 filamenti.
La monellina, una cistatina dal sapore dolce proveniente dalla bacca tropicale *Dioscoreophyllum cumminsii*, rappresenta il primo caso identificato di proteina fibrillogenica da pianta. Delle 2 catene polipeptidiche da cui essa è formata, soltanto una, la catena B, è in grado di formare fibrille in condizioni destabilizzanti.
Ho analizzato la dinamica conformazionale, la stabilità termica e il potenziale fibrillogenico di due cistatine da piante. La prima, *mnei*, è una monellina ricombinante a singola catena polipeptidica, ottenuta collegando con un dipeptide le due catene da cui è costituita la monellina *wild type*; la seconda, *oryc*, è la cistatina I dalla variante di riso *Oryza sativa,* la prima cistatina da pianta che è stata ben caratterizzata.
I risultati ottenuti con esperimenti di denaturazione termica hanno mostrato che entrambe le proteine sono molto stabili anche a pH acido e subiscono un processo di denaturazione termica a due stadi, con assenza di intermedi. Tramite analisi di spettroscopia in dicroismo circolare abbiamo dimostrato che entrambe le proteine in ambiente acido conservano la tipica struttura delle cistatine. Tuttavia, incrementando la temperatura, è stata identificata una transizione dalla struttura β-sheet verso una struttura *random coil* .
Nel caso della proteina mnei, prolungando l'esposizione in condizioni acide, è stata evidenziata una ulteriore transizione verso una struttura β, il cui segnale viene perso con il progredire dell'incubazione, ciò che indica la formazione di aggregati insolubili. La cinetica di formazione di tali aggregati è stata seguita con il fluoroforo Tioflavina T, specifico per le strutture amiliodi. Dopo circa 5 ore di incubazione in ambiente acido, è stato possibile identificare aggregati fibrillari con la tipica struttura amiloide β-*cross*, capaci di legare il fluoroforo.
Durante l'incubazione della cistatina mnei in ambiente acido sono state inoltre effettuate analisi per gel-filtrazione, volte a caratterizzare gli aggregati generati. Attraverso tali analisi è stato possibile rivelare la progressiva scomparsa nel tempo della forma monomerica solubile della proteina, plausibilmente coinvolta nel processo di polimerizzazione fibrillogenica. Sono state infine condotte analisi preliminari di microscopia a forza atomica su un campione proteico dopo incubazione. Da tali analisi è stata evidenziata la presenza di materiale pre-fibrillare oltre a quella di aggregati amorfi.
Analoghi esperimenti sono stati effettuati sulla proteina oryc. Gli aggregati di oryc non generano la tipica struttura amiloide fibrillare, come si deduce dall'assenza di legame della Tioflavina T. Durante l'incubazione della proteina in ambiente acido sono stati condotti esperimenti per gel-filtrazione, con lo scopo di analizzare le specie molecolari coinvolte nel processo. In seguito a tali analisi, dopo circa 3 ore di incubazione, è stata isolata una forma oligomerica della proteina, dal peso molecolare corrispondente ad un trimero. Tale specie si forma a partire da unità monomeriche, come si evince dal fatto che la quantità di specie monomeriche diminuisce nel tempo, parallelamente all'aumento della specie trimerica.
Si è ipotizzato che la assenza di polimeri amiloidi durante l'incubazione di oryc sia dovuta alla formazione dell'oligomero stabile isolato mediante gel-filtrazione. Questo

potrebbe sequestrare il monomero impedendone la polimerizzazione in strutture fibrillari.

Infine, il mio progetto di ricerca ha previsto uno studio sulle potenziali applicazioni biotecnologiche dei polimeri fibrillari. Infatti, le fibrille amiloidi rappresentano un esempio di *molecular self-assembly* per la costruzione di strutture polimeriche ordinate. Il *molecular self-assembly* opera attraverso il riconoscimento di superfici molecolari complementari, che determina una loro stabile associazione. Esso viene utilizzato nella costruzione di nuovi materiali attraverso la strategia del *bottom-up,* che opera su singole molecole per costruire strutture supermolecolari. Il *molecular self-assembly* è un meccanismo comunemente utilizzato nel campo delle nanobiotecnologie.
La struttura amiloide rappresenta un elemento di interesse nel campo dei nuovi materiali, per le sue specifiche proprietà. E' un polimero non covalente, periodico e ordinato e rappresenta una struttura resistente che può svilupparsi nelle tre dimensioni.
In questo lavoro ho esplorato la possibilità di generare nuove matrici catalitiche utilizzando le fibrille amiloidi come supporto a cui legare un enzima di interesse biotecnologico. Il termine "fibrille catalitiche" è apparso di recente nell'ambito delle nanotecnologie.
Per costruire fibrille catalitiche, abbiamo utilizzato una proteina di fusione come unità base del processo di fibrillogenesi. La proteina di fusione è costituita dall'enzima glutatione S-transferasi (GST), che rappresenta la porzione catalitica, e dal polipeptide [1-93]ApoA-I, che rappresenta la porzione fibrillogenica. La scelta dell'enzima GST è giustificata dal fatto che questo enzima è ampiamente utilizzato in ambito biotecnologico, in particolar modo nei processi di detossificazione e di *bioremediation*. La proteina di fusione, definita GST-[1-93]ApoA-I, è stata espressa in un opportuno ospite procariotico e successivamente isolata all'omogeneità mediante cromatografia di affinità e cromatografia a scambio ionico.
Per poter realizzare fibrille catalitiche devono essere soddisfatti due requisiti, e cioè che l'attività enzimatica della componente catalitica sia preservata durante l'aggregazione in fibrille e che la presenza della componente catalitica non ostacoli il processo di aggregazione in fibrille. Pertanto il prodotto di fusione, una volta isolato, è stato sottoposto ad analisi volte a definire l'attività catalitica e il potenziale fibrillogenico.
La determinazione dell'attività catalitica mediante il saggio del CDNB (1-chloro-2,4-dinitrobenzene) ha indicato che la proteina di fusione solubile mantiene invariata l'attività dell'enzima GST. Inoltre, l'analisi della struttura della proteina di fusione, effettuata mediante dicroismo circolare a pH 8.0, ha fornito uno spettro sovrapponibile a quello ottenuto con l'enzima GST. Il lieve aumento riscontrato in relazione alla struttura *random coil* della proteina di fusione è plausibilmente da addebitare alla presenza del polipeptide destrutturato [1-93]ApoA-I.
E' stato quindi analizzata la capacità della proteina di fusione a generare fibrille utilizzando approcci sperimentali indipendenti. In seguito a debole acidificazione del valore di pH, si è verificato che la proteina di fusione è in grado di legare il fluoroforo ANS, evento da correlare ad una transizione conformazionale, indotta dal pH, verso un intermedio parzialmente strutturato a carattere di *molten globule*. Tale transizione è risultata reversibile se il valore di pH era riportato ad 8.0.
Per verificare che la proteina di fusione fosse in grado di generare tipiche fibrille amiloidi, abbiamo condotto saggi spettrofluorimetrici e spettrofotometrici utilizzando

coloranti specifici per strutture amiloidi, quali Tioflavina T e Congo Rosso. I risultati ottenuti hanno indicato che, dopo circa 40 ore di incubazione in condizioni che inducono l'aggregazione, sono presenti tipici aggregati amiloidi generati dalla proteina di fusione. In parallelo, il materiale insolubile generato dall'incubazione dell'enzima GST è risultato invece negativo alla presenza di strutture fibrillari. Infine, le analisi per microscopia a forza atomica, condotte dal gruppo della Prof. A. Gliozzi, Università di Genova, hanno dimostrato la presenza di filamenti fibrillari generati dalla proteina di fusione.

Il materiale fibrillare, immobilizzato su filtri di PVDF con *cut off* pari a 0.02 o 0.2 μm, è stato sottoposto ad analisi di attività enzimatica, facendo fluire in continuo attraverso i filtri una miscela contenente i substrati della reazione dell'enzima GST. Tale procedura ha permesso di dimostrare che il materiale fibrillare, immobilizzato sulla membrana con *cut off* di 0.02 μm, era cataliticamente attivo, cioè in grado di generare il prodotto enzimatico atteso.

INTRODUCTION

1. The protein folding problem

Protein folding is crucial in completing the information transfer from DNA to an active protein product [1]. In 1973 C. Anfinsen stated a fundamental principle of the protein folding, i.e. the folding of a protein is determined by its amino acid sequence: *" All the information necessary to achieve the native conformation of a protein in a given environment is contained in its amino acid sequence"* [2].

From this principle derived the structure-function paradigm, claiming that a specific function of a protein is determined by its unique and rigid three-dimensional structure. However, since from those years, some questions emerged, one of these being the Levinthal paradox [3], in consideration of the short time required for protein folding *in vivo* as well as *in vitro* with respect to the timescale for a random-search mechanism for protein folding.

More recently, "natively unfolded" proteins emerged [4] as proteins lacking of almost any secondary structure and shown to be extremely flexible and disordered under physiological conditions [5]. This proteins are also called "natively denatured" [6], "intrinsically unstructured" [7] or "intrinsically disordered" [8]. In 2000, more than 100 proteins were classified as natively unfolded [9], this number being actually greater. The main feature of these proteins is the intrinsic structural plasticity, meaning that they can undergo a disorder to order transition upon functioning [10,11,7].

Finally, in spite of the great number of proteins in all known organisms (10^{11}), through the exploration of the Protein Data Bank (PDB) less than 700 different folds have been observed to date and the proteins domains exhibit only 32 types of different architectures [13].

These evidence led to a reconsideration of the protein folding problem. Even if more folding models were proposed in the last 30 years, actually the most reliable one is the energy landscape, funnel-like model, proposed by P.G. Wolynes: *"to fold, a protein navigates with remarkable ease through a complicated landscape"* [14].

The so called "new view" of the protein folding shows that an ensemble of unfolded polypeptide chains search for a predominantly native state on a usually rugged energy surface, or 'landscape', through a transient ensemble of partially folded species (Fig .1). This model presents some advantages:

1. It can be adapted to different proteins. For small proteins this landscape appears to be smoother and the folding is very fast and follows a two state pathway, with only two species being stably populated , i.e. the ensemble of unfolded structure and the native state. In contrast, larger proteins have a rougher energy landscape, which indicates a slower and multistate folding that allows the population of partially folded species. These species may be *on-pathway* if they are productive for the correct folding or *off-pathway* if they are traps that inhibit the protein folding.
2. It takes in account both the kinetic and thermodynamic aspects of the folding pathway. The folding rate is slowed by ripples in the energy landscape corresponding to local minima populated by transiently stable intermediates. Kinetic traps arise from energy barriers; when barriers are high enough the protein is trapped. By contrast, the steeper the slope, the faster the folding. Towards the bottom of the folding funnel, the number of protein conformations decreases, as does the chain entropy. According to the model, there are parallel micropathways that form microstructures in several region of the

protein reducing the number of possible conformation. The Levinthal paradox is so removed.

3. It provides the way to understand the formation of stable species different from the native structure, like aggregates, oligomers or fibrils. A competition between intramolecular, native-like interactions and intermolecular, non native interactions occurs generating fluctuations in the unfolded or partially folded ensemble. The former drive the reaction towards the native structure, the latter drive the reaction towards other stable structures.

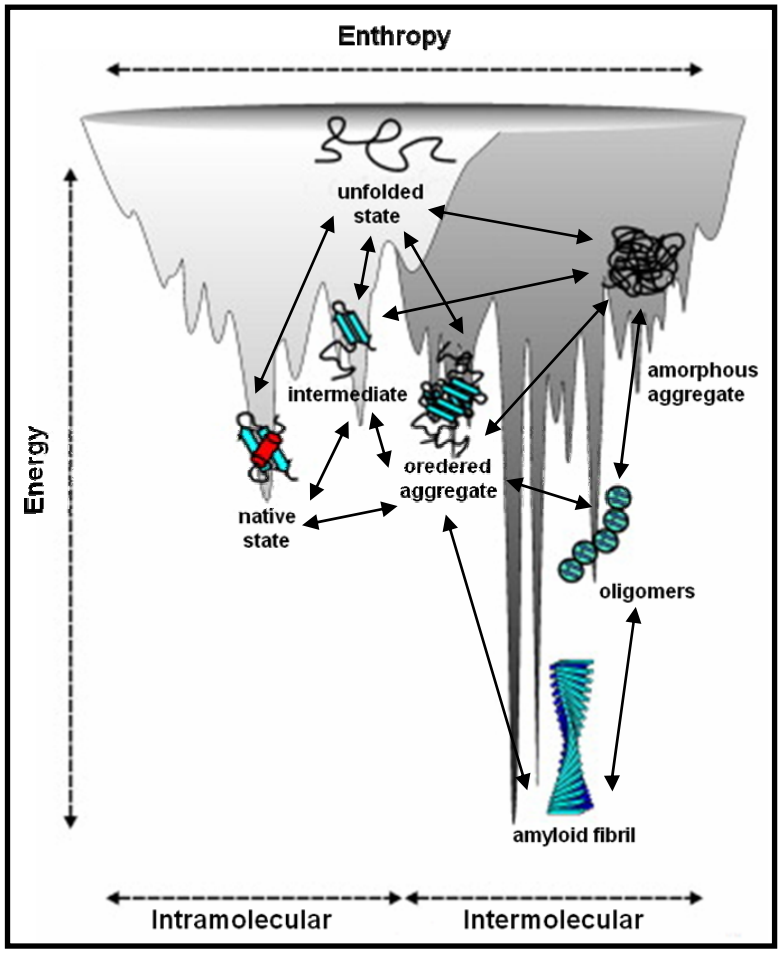


Fig. 1: Energy landscape for protein folding and aggregation. The surface illustrates the roughness of the protein energy landscape, showing the multitude of conformational states available to a polypeptide chain. Rather simple folding funnels (light grey) can describe the conformational search of a single polypeptide chain to a functional monomer by means of intramolecular interaction. Intermolecular protein association dramatically increases ruggedness (dark grey). Arrows indicate the proposed pathways linking the conformational states represented in figure.

By means of NMR methods and X-ray crystallography, much is known about the structure of proteins in their native conformations. Aggregates and amyloid fibrils are actually under studies as crucial in the emerging diseases called amyloidosis (see

later). On the other hand, the conformational properties of unfolded proteins and partially folded states are more difficult to define because of their heterogeneity, complexity and rapid interconversion.
Early intermediates belonging to the ensemble of partially folded states are the hydrophobic collapsed state, the pre-molten globule and the molten globule state, occurring in this order in the pathway towards the folded state. The time scale of the appearance of these intermediates range between nanosec to millisec, making difficult the analyses of the initial events of the folding, as they occur within the dead time of the most common analytical device.
Better characterized is the molten globule state, a general intermediate in the folding pathway of proteins. It is a compact intermediate with a high content of native secondary structure [15, 16], but a fluctuating tertiary structure. It contains accessible hydrophobic surfaces which bind a hydrophobic dye, anilinonaphtalene sulphonate (ANS); this behavior should be considered as a characteristic property of the molten globule state [17,18]. Finally the molten globule state appears as a heterogeneous species, in which the helical domain is structured, hydrophobic interactions are fluctuating and the β sheet is rather disordered.
In the final rate-limiting step, the protein achieves its native conformation with the emergence of functional properties. These final events correspond to the precise ordering of the elements of secondary structure, the correct packing of the hydrophobic core, the correct domain pairing in multidomain proteins, the reshuffling of disulfide bonds by the protein disulfide isomerase, the cis-trans proline isomerization by peptidyl-prolyl-cis-trans isomerase and the subunit assembly in oligomeric proteins. For several proteins, the rate-limiting step requires the reorganization of misfolded species [19]
A multidisciplinary approach is commonly used to investigate the protein folding, encompassing biophysic, biochemistry, biology, medicine and mathematics.
X-ray crystallography, by means of missed electron density [20] and NMR techniques, including solid state NMR, has been successfully used to analyze several unfolded, partially folded and aggregated proteins [7]. Near UV CD and Far-UV CD spectra are characteristic of protein tertiary and secondary structures [21] and can be used to determine conformational changes or aggregation [22, 23]. Protein denaturation may be easily detected with CD analyses too. Techniques such as gel-filtration, sedimentation, and dynamic and static light scattering may help in defining the compactness of a protein [9]. Additional knowledge on the compactness of a protein structures may be extracted from the analysis of different fluorescence or colorimetric characteristics. Increased proteolytic degradation *in vitro* can indirectly confirm an increased protein flexibility or accessibility [8]. Immunochemical methods may also be applied toward the elucidation of protein structural changes [24] and aggregation [25]. Overall, these techniques add important information to the conformational description of a polypeptide.
These are the most common techniques used in protein folding characterization, even if advances in these methodology have been recently made, not reported here.
Finally, it has to be considered that protein folding occurs *in vivo* where the complexity of the pathway drastically increases, as phenomena like protein crowding or confining act on the protein folding in a manner not always predictable. Moreover, molecular chaperones have a role in preventing improper folding and subsequent aggregation, while degrading systems, like the ubiqitin-dependent proteasome, or the ubiquitin-independent autophagy, remove proteins with an incorrect folding.

2. Protein aggregation and diseases.

Protein aggregation is a side reaction along the folding pathway that can generate oligomers, disordered aggregates, i.e. amorphous aggregate, or highly ordered aggregates, i.e. amyloid fibrils, as shown in Fig. 1.

The generally accepted hypothesis for protein aggregation is that either from partial structuring of unfolded polypeptide (*refolding* model), or from limited conformational changes in native structure (*gain of interaction* model), a set of partially folded intermediates is populated [26] (Fig. 2). In these intermediates aggregation competent regions previously buried are exposed, thus permitting aggregation. In the initial stages of aggregation, surface elements of one molecule interact with matching surface areas of neighbouring molecules [27] Even if the major driving forces, i.e. the formation of hydrogen bonds and the burial of hydrophobic surface area, are commonly involved in both protein folding and aggregation, specific residues are involved [28] and intermolecular interactions promote aggregation, while intramolecular interactions promote protein folding [49].

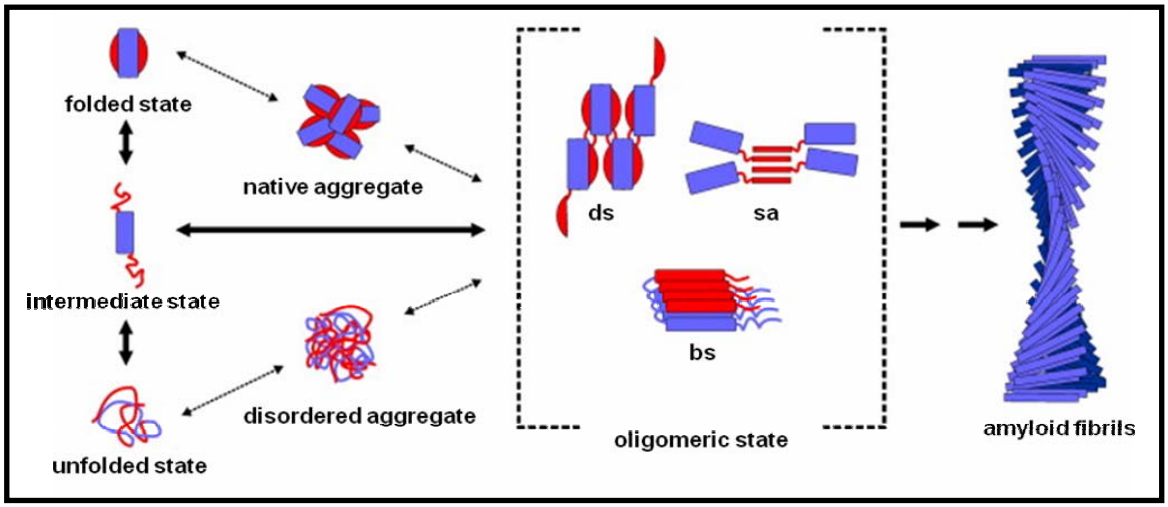


Fig. 2. Proposed mechanisms of protein aggregation. Amyloid fibril formation proceeds from intermediate partially folded states that are formed via partial unfolding of the native structure, or via the partial structuring of unfolded polypeptides. Ordered aggregates associate via mechanisms such as domain swapping (ds), strand association (sa) or β-strand stacking (bs). Self-association of these early oligomeric species, possibly involving further conformational changes, leads to the formation of amyloid fibrils.

Amyloid fibrils are formed with nucleation-dependent polymerisation, in which the protein monomer is converted into a fibrillar structure *via* a transiently populated aggregation nucleus [29]. Following the rate-limiting step of nucleus formation, which results in the presence of a typical lag-phase, aggregate growth proceeds rapidly by further addition of monomers, or other assembly-competent oligomers, to form larger polymers. Growing polymers are commonly called protofibrils.

The formation of protein oligomers requires a critical concentration of proteins [29, 30] Basing on kinetic modelling of protein aggregation, the early protein oligomers act as seeds facilitating protein misfolding and promoting protein polymerization and

eventually fibrils formation [31]. In the slow lag phase, a series of kinetically unfavourable interactions occurs to form the oligomeric nucleus that rapidly grows to form larger polymers. In the nucleation dependent polymerization model the lag phase can be shortened in the presence of small amount of preformed fibril seeds. According to the recycling hypothesis, an equilibrium is reached on the growing surface of the fibrils between associating and dissociating molecules as in a reversible reaction [32].
As shown in Fig. 2, three main mechanisms are proposed for amyloid fibril formation. In the strand association mechanism, the amyloid fold occurs *via* the direct stacking of single protein subunits [33]. By contrast, in the domain swapping model an exchange of a specific domain between the assembling subunits is required, but the main protein fold remains the same [35, 36]. In the generic β–cross spine mechanism [34], the fibrils grow by providing short polypeptide segments that stack together to form the stable fibril spine. However, in all proposed mechanisms a specific protein destabilisation is required, that results in an increased population of the partially folded conformations.
In vitro destabilization depends from protein mutation and intrinsic instability [57, 58], protein concentration, pH [38], temperature [37], ionic strength, metals or redox environments [39], co-solute or denaturants.
In vivo destabilization depends also from the presence of abnormal proteolytic cleavage [40], the impairing of the cell proteasome [41] and the role of chaperones [42], the protein crowding and confining [43], the changes in local environment, the altered post translational modification or genetic defects [40, 57]. The late onset of aggregation related diseases (see later) suggests a crucial role of cellular mechanisms for removing the aggregates, which becomes ineffective or overwhelmed with the age.
In spite of the difficulties to obtain high-resolution structures of amyloid fibrils, as these species are insoluble and non-crystalline, the typical amyloid structure is defined in terms of empirical observations from X-ray fibre diffraction, EM/AFM and specific chemical staining [44, 45].
The first definition of amyloid as possessing a generic cross-β structure was made nearly 40 years ago [46]. EM and AFM studies have shown that amyloid fibrils are straight, unbranched and about 70–120 Å in diameter [47]. They are composed of several protofilaments, each protofilament consisting of contiguous β-pleated-sheet polypeptide chains, perpendicularly aligned to the long axes of the fibers (cross β–pleated sheets). Fibrils of different diameters and morphologies have been observed, related to the so called amyloid fibrils polymorphism [56, 70]. Furthermore, green birefringence upon staining with Congo red, or the fluorescence shift after staining with ThT, are also classical features of the amyloid fold [48]. However, other techniques are now commonly used to detect amyloid aggregates, like ultracentrifuge, light scattering or gel-filtration chromatography.
An increasing number of human diseases has been linked to protein aggregation and to the aberrant accumulation of protein deposits in different tissues and organs [49, 50]. These disorders are collectively known as *protein misfolding* or *conformational* diseases [51, 52], because the pathological protein undergoes structural changes that results in self-association, aggregation and tissue deposition. The pathological proteins are called *chameleon* protein indicating their conformational plasticity [53]. When protein aggregates result in ordered, insoluble, mainly extracellular deposits, the related diseases are called amyloidoses. Rudolph Virchow in 1854 adopted the term "amyloid", first introduced by Schleiden in 1838 to describe plant starch, in

referring to tissue deposits of material that stained in a similarly to cellulose when exposed to iodine.
In amyloid diseases the protein deposits contain fibrillar protein assemblies characterized by the specific dye-binding and tinctorial properties, the cross-β X-ray fibre diffraction pattern and macroscopic long, straight and unbranched morphology previously described. Deposits can be almost undetectable, or reach kilograms of deposited protein [54] and always contain carbohydrates and other proteins. Depending on the localization of the deposits, amyloidoses can be systemic or localized [55].
Conformational diseases include Alzheimer's, Parkinson's and Huntington diseases, spongiform encephalopathies (or prion encephalopathies), ataxias, etc. All these diseases have a high impact on public health, considering the late onset and the degenerative trait.
There are more than 20 human amyloidogenic proteins known to abnormally self-assemble into fibrils of variable length [59]. They are all characterized by a β-cross repeat structure [60]. Table I summarizes the amyloidogenic proteins and the corresponding pathologies.

Clinical syndrome	Precursor protein	Fibril component
Alzheimer's disease	Amyloid protein precursor	Aβ-Peptide 1-40 to 1-43
Primary systemic amyloidosis	Immunoglobulin light chain	Intact light chain or fragments
Secondary systemic amyloidosis	Serum amyloid A	Amyloid A (76-residue fragment)
Senile systemic amyloidosis	Transthyretin	Transthyretin or fragments
Hereditary cerebral amyloid angiopathy	Cystatin C	Cystatin C minus 10 residues
Hemodialysis-related amyloidosis	β2-Microglobulin	β2-Microglobulin
Familial amyloid polyneuropathy III	Apolipoprotein A-I	N-terminal fragments of apolipoprotein A-I
Type II diabetes	Islet amyloid polypeptide (IAPP)	Fragment of IAPP
Spongiform encephalopathies	Prion	Prion or fragments thereof
Hereditary non-neuropathic systemic amyloidosis	Lysozyme	Lysozyme or fragments thereof
Injection-localized amyloidosis	Insulin	Insulin
Hereditary renal amyloidosis	Fibrinogen	Fibrinogen fragments

Table I: The most common amyloidoses, the corresponding pathogenic protein precursor and their components detected in the amyloid fibrils.

Fig. 3 shows some proteins able to form amyloid fibrils, ranging from natively unfolded polypeptides, through polypeptides possessing extensive α-helical

structure, to proteins containing β-sheet structure. Aggregation prone stretches are also indicated, identified using the most common algorithm for predicting propensity of proteins to aggregate, the TANGO algorithm [62]. The extreme diversity in structure and function of proteins associated with amyloid fibrils in human diseases, together with the observation that many proteins (i.e. RNase A or mioglobin) not related to amyloidoses are able to form *in vitro* fibrils, suggests that the amyloid fold is an ancestral motive preserved during the evolution, a universal, inherent property of all polypeptide chains related to the global free-energy minimum (the thermodynamic ground state of the aggregation).

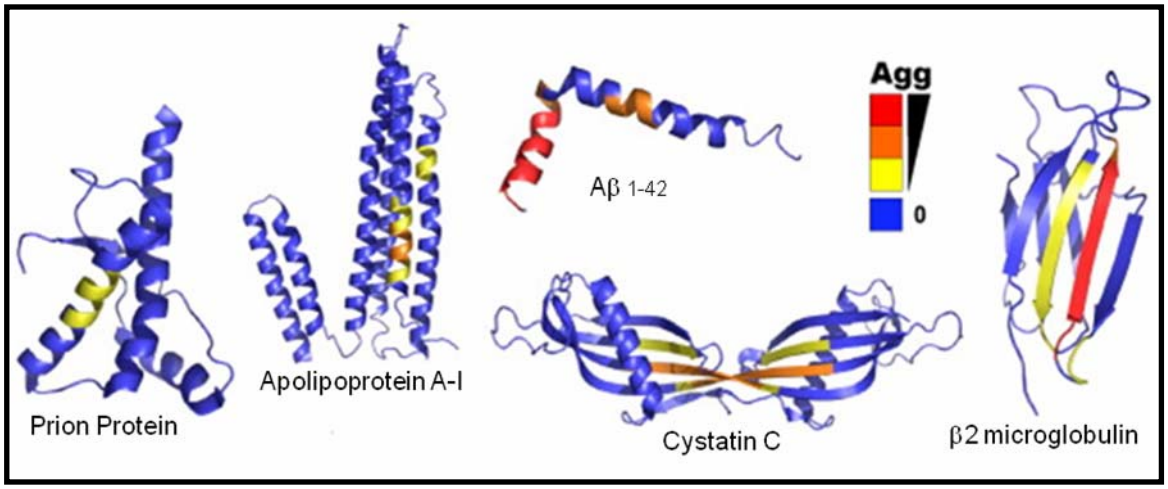


Fig. 3. Representative structures of proteins involved in disease-related amyloid fibril formation. The coloured regions are polypeptides with high aggregation propensity, predicted according to the Tango algorithm. Sequences shown in blue are predicted to have no β-aggregation propensity, while polypeptide stretches coloured in yellow, orange and red indicate an increasing propensity to aggregate. Notably, the proteins shown are different in structure and function.

Two main reasons may explain the persistence of the amyloid folding motif. It has been hypothesized that has a functional active role [61], or, alternatively, that it protects cells toward cytotoxic misfolded proteins or early aggregates [63]. In fact, in bacteria [66, 67], fungi [64, 69] insects [68] and also in humans [65] different processes have been related to the amyloid structure, like structure maintenance, information transfer, human hemostasis and mammals melanin biosynthesis. With respect to the supposed protective role of amyloid structures, amyloid fibrils are thought to be not cytotoxic, in contrast with early aggregates occurring during protein misfolding and having inherent toxicity [63, 104, 106].
Even if there is no a clear relationship between protein sequences and aggregation, proteins able to aggregate are characterized by a low sequence complexity [71] and/or high net charge coupled with low mean hydrophobicity [72] and by the abundance of residues favouring the β-sheet secondary structure [74]. By contrast, proline or glycine residues [73, 75], edge strands in native β-sheets structure [76], or a long stretch of alternating polar and non polar amino acids [77] are thought to be negative determinants in amyloid fibril formation and are commonly called “structural gatekeepers” [78].

One of the most interesting features of conformational diseases is the mechanism of infectivity. S. Prusiner in 1982, while studying the prion disease, established the "protein only" concept, i.e. the existence of a "*proteinaceous infectious particle that lacks nucleic acid*" [79]. Indeed, no nucleic acid has been found to be involved in the propagation of the infectious agent. Prion diseases may result from genetic, infectious or sporadic disorders, but in each case an aberrant conformational change in a protein is propagated. The self-replication of conformational information enables prions to create a conformational memory [81] and to transmit the disease [80] even if only in a species-specific manner. Prion particles with the same polypeptide sequence can have distinct physiological effects associated to different conformational states of prion structure [82, 83].

3. Apolipoprotein A-I and amyloidosis.

Human apolipoprotein A-I (ApoA-I) represents the major structural component of the serum high density lipoproteins (HDL), having the ability to avidly bind lipids [84]. ApoA-I plays a critical role in lipid metabolism [85], both in delivering cholesterol to steroidogenic tissues and in transporting it from the periphery to the liver for catabolism, in the so called reverse cholesterol transport (RCT). Because of its active role in RCT, it exerts an anti-atherogenic activity *in vivo* protecting against cardiovascular diseases [86].

The liver and the intestine synthesize ApoA-I; about 5-10% of human plasma ApoA-I exists in a lipid free state. By contrast, ApoA-I represents roughly 70% of the HDL protein mass. ApoA-I gene is located on the long arm of chromosome 11. The protein is synthesized as a pre-pro-protein of 243 amino acids. After the cleavage of the 18-residue signal polypeptide, the pro-protein is secreted in the plasma, where the 6-residue pro-peptide is removed by cleavage by a still unidentified metalloprotease [87]. The mature, circulating ApoA-I protein consists of a single polypeptide with a 28 kDa mass. Transcriptional regulation of ApoA-I gene is complex, involving induction by several hormones [88] and inhibition by anti-oxidant molecules [89]. No post translational modifications, like glycosilation or disulfide bonds formation, occur.

Upon binding to lipids, the secreted lipid free ApoA-I generates the nascent, discoidal HDL. The conversion of unesterified cholesterol into cholesteryl ester by the enzyme lecithin:cholesterol acyltransferase (LCAT) converts the nascent, discoidal HDL into mature spherical HDL.

During the lipid transport, ApoA-I adopts different lipid association states and interacts with several receptors, like the ATP binding cassette transporter (ABCA1) or the scravenger receptor (SR-B1), as well as with enzymes, like LCAT, cholesteryl ester transfer protein (CETP) or phospholipid transfer protein (PLTP) This suggests a great conformational plasticity of the protein, commonly included in the natively unfolded protein group [105].

The protein likely exists as an ensemble of related and interconverting species, including a molten globule state lacking a unique defined tertiary structure [90]. ApoA-I catabolism is complex as it requires HDL dissociation. The kidney is the principal site of ApoA-I degradation [91].

A comparative analysis of the primary structure of ApoA-I of mammals shows that the N-terminal region is highly conserved [92]. Secondary structure analyses show that the protein is organized into eight α-helical segments of 22 amino acids and two 11-mer repeats that are frequently separated by proline residues [93]. (Fig 4A). These helices are predicted to be amphipathic with a hydrophobic face that likely mediates lipid interactions and a polar face that interacts with water. By contrast, the N-

terminal region contains the most ambiguously defined structure of the protein, being this region highly flexible.
To date, two crystal structure of ApoA-I are deposited, showing very different structures. The full-length lipid free ApoA-I was crystallized by Ajees et al. [94] (PDB 2A01) The structure (Fig. 4B) indicates that the N-terminal two-thirds of the molecule is involved in an intramolecular four-helix bundle organization, while the 50 residue C-terminal region forms an independent hairpin domain.
Borhani et al. [95] (PDB 1AV1) crystallized a deletion mutant lacking the N-terminal 43 amino acids. The structure (Fig. 4C) shows a ring-shaped assembly of four ApoA-I molecules arranged in extended, kinked α-helices. However, because of the missing N-terminus and the tetramerisation, the latter structure appeared to be more applicable to lipid-bound ApoA-I, while the former structure is more consistent with a lipid-free ApoA-I.

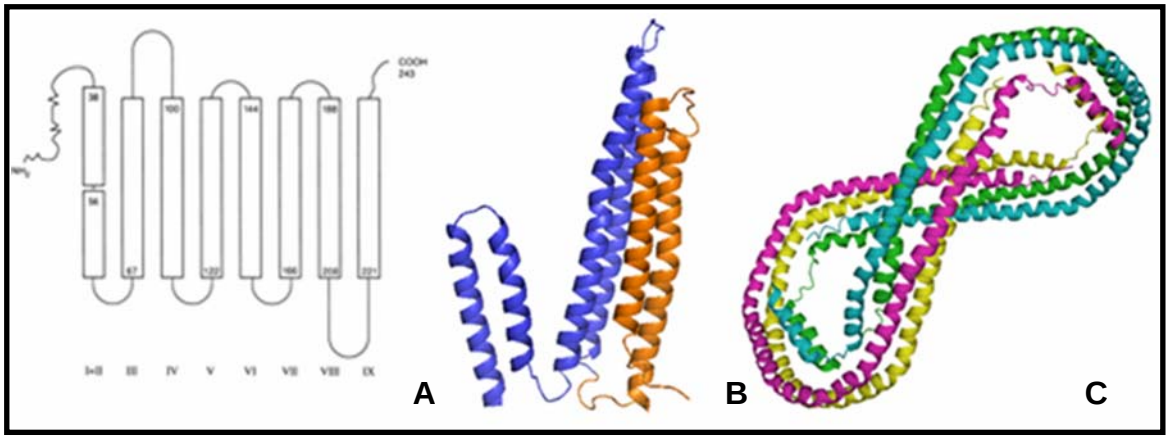


Fig. 4: (A), ApoA-I secondary structure prediction; (B), lipid free ApoA-I crystal structure (in red the N-terminal region); (C), the crystal structure of ApoA-I proposed for the lipid-bound state.

More than 50 naturally occurring ApoA-I variants have been reported, half of them associated with HDL reduced levels [96]. Among them, ApoA-I_{MILANO} variant exerts a protective role against atherosclerosis [97].
Several variants of ApoA-I are associated to amyloid diseases [84]. These variants aggregate *in vivo* and form amyloid fibril protein deposits with the typical β-cross structure [70, 84]. Fibrils accumulate in tissues and organs, causing severe consequences such as heart, renal or liver failure [98] (Fig. 5A). The complete list of the 13 ApoA-I amyloidogenic mutations is reported in Fig. 5B and Table II. In the latter, the affected organs or tissues are indicated for each mutant.
The main trait of all ApoA-I amyloidoses is an autosomic dominant transmission. All patients have single heterozygous mutations.
Most of the mutants are generated by single point mutations in the 243-amino acid sequence of the native protein, although deletions or deletions/insertions have also been described [57, 99, 100]. Some of the single point mutations add a polar or a positively charged amino acid, while other mutations add a proline residue.
In ApoA-I variants, the mutation is either positioned inside the N-terminal portion of the protein that is eventually found in fibrils (internal mutations), or it occurs in a position located externally to the fibrillogenic polypeptide (external mutations).
Analysis of the amyloid fibrils extracted from patients affected by ApoA-I derived

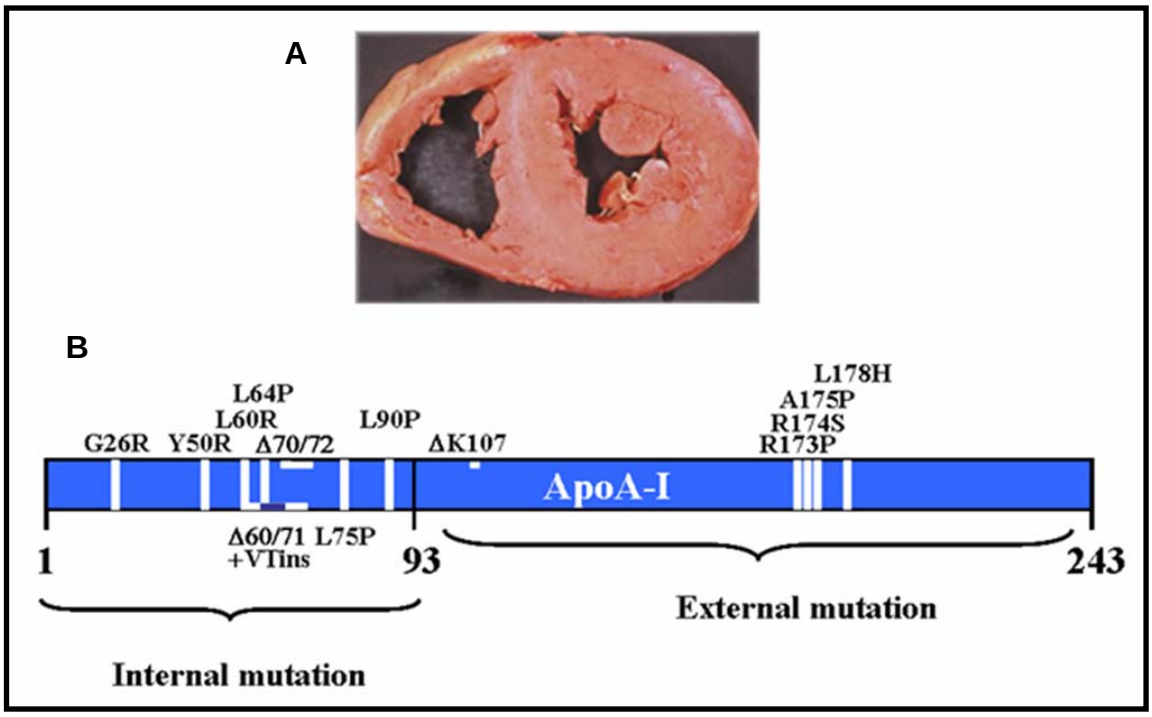


Fig. 5: (A), Increased dimension heart section from a transplanted patient affected by ApoA-I amyloidosis. (B), A map of ApoA-I amyloidogenic "internal" and "external" mutations. Single point mutations (vertical white bars), insertions (horizontal white bars) or deletions (horizontal blue bars) are indicated.

Mutant	Organ or Tissue	Ethnic origin
Gly26Arg	Nervous system, Kidney	USA, UK, Scandinavy, Ireland
Trp50Arg	Liver, Kidney	UK
Leu60Arg	Kidney	UK
Δ60/71 insVal/Thr	Liver	Spain
Leu64Pro	Kidney	n.s.
Δ70-72	Kidney	South Africa
Leu75Pro	Liver, Kidney	Italy, USA
Leu90Pro	Heart, Skin	France
ΔLys107		n.s.
Arg173Pro	Heart, Larynges, Skin	USA, Caucaso
Leu174Ser	Heart	Italy
Ala175Pro		n.s.
Leu178His	Heart, Larynges, Skin	France

Table II: The mutations occurring in ApoA-I sequence, associated to amyloid pathology. The affected organs or tissues and the ethnic origins are also indicated.

amyloidosis has shown that natural fibrils consist of ApoA-I N-terminal fragments 80-100 residues long.
For both external and internal mutations the about 10 kDa N-terminal ApoA-I polypeptide represents the main protein species in amyloid fibrils.
In particular, the ApoA-I variant carrying the substitution of leucine at position 174 with serine (L174S) [101] is associated with a severe hereditary systemic amyloidosis predominantly involving the heart. Amyloid fibrils, extracted from a heart transplanted patient were found to be mainly constituted by the amino terminal polypeptide (residues 1-93) of ApoA-I [101, 149]. The point mutation located at position 174 of full length ApoA-I renders this protein prone to be cleaved by an unknown protease, releasing the amino terminal polypeptide.
Polypeptide 1-93 extracted from *ex vivo* amyloid deposits was found to be largely unstructured at neutral pH, but intrinsically amyloidogenic and able to form aggregates and fibrils upon acidification [102]. A complex fibrillogenic pathway, in which the fibrillogenic polypeptide assumes a random coil structure at neutral pH, shifts into an unstable helical conformation at acidic pH and then aggregates in to a β-sheet-based polymeric structure, was envisaged [102]. Interaction of the amyloidogenic polypeptide with phospholipids prevents the switch from the unstable helical conformation to the β–sheet form. Moreover, by atomic force microscopy an ultrastructural description was performed on the natural amyloid material extracted from the heart of a transplanted patient, showing the presence of both globular aggregates and fibrils [103]. The fibrils appeared classically twisted and composed by three protofilaments.
Nevertheless, the possibility to elucidate the complex and intriguing fibrillogenic pathway of ApoA-I amyloidogenic polypeptide had been limited so far, due to the low amount available from in vivo sources.
Recently, in the laboratories of Prof. Piccoli of the University of Naples, an effective and reliable expression system has been settled to produce a recombinant version of 1-93 polypeptide as a pure and stable product, suitable for conformational and fibrillogenic studies.

4. Cystatins and amyloidosis

Cystatins are a family of small proteins whose main characteristic is that of being a cysteine protease inhibitors [107]. Cysteine protease inhibitors are widely distributed in mammals, plants, and insects and may function to protect cells from unwanted proteolysis and to control intra- and extracellular protein breakdown. They are acidic proteins typically composed of 100-120 amino acids, with 4 conserved cysteine residues that form 2 disulphide bonds [108].
During the 1980s, several sequences with high omology to cystatins were discovered. Although some of them are not endowed with inhibitory properties and do not contain disulphide bridges, they were collectively classified, together with cystatins, in the cystatin “superfamily” [109]. The cystatin superfamily consists of tightly and reversibly binding inhibitors of cysteine proteases of the papain family. It in higher animals can be divided into three types, based on the molecular mass and the presence of disulfide bonds (1-3). Type I animal cystatins, about 100 amino acid long, lack disulfide bonds, and are present predominantly in the intracellular compartment. Type II cystatins, about 120 amino acid long, contain at least two characteristic disulfide bonds and are found extracellularly, in body fluids. Type III cystatins are larger glycoproteins containing several type 2-like cystatin domains.

Thus far, the crystal structure of several cystatins, including human cystatin B (also referred to as stefin B) (PDB 1stf) [110] and human cystatin A (also referred to as stefin A) (PDB 1dvc) [111, 112] have been resolved. These structures have shown that cystatins share a common typical fold, consisting of a five-stranded antiparallel β–sheet structure that wraps around a central α-helix (Fig. 6A).

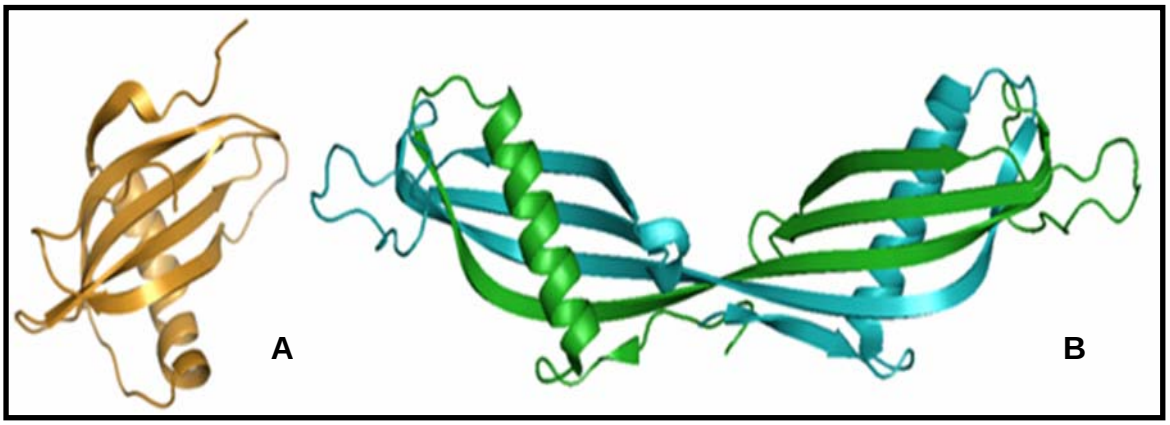


Fig. 6: (A) Cystatin A (PDB 1dvc) structure showing the common typical fold consisting of a five-stranded antiparallel β-sheet structure that wraps around a central α–helix. (B) 3D domain swapped cystatin C (PDB 1tij). The two chains are in green and blue respectively.

Some cystatins are involved in human amylopathies. Amyloid aggregates of human cystatin C are found, together with the amyloid Aβ polypeptide, in brain arteries of old people affected by cerebral amyloid angiopathy [113, 114]. In addition, human cystatin C plays an important role in the development of cerebral hemorrhage in people with hereditary cystatin c amyloid angiopaty (HCCAA), an autosomal dominant desease characterized by the accumulation of amyloid fibrils in many tissues, including cerebral arteries [115]. Patients affected by HCCAA harbor a mutated version of cystatin C, with leucine 68 mutated to a glutamine, that gives rise to deposits of amyloid fibrils in the cerebral arteries, resulting in brain hemorrhage and death in young adults [116]. Recently, loss of function mutation in the gene encoding human cystatin B, has been suggested as a possible cause of a severe neurological disorder known as the Unverricht-Lundborg disease [117].
For human cystatin C, a detailed mechanism of amyloid fibril formation has been proposed. The solid state structure of wild type cystatin C [118] (PDB 1g96) is a very stable dimer originating from 3D domain swapping [119] (Fig. 6B). This observation led Janowsky et al. to suggest also a three dimensional domain swapping process to explain the formation of cystatin C amyloid fibrils [118, 120]. Domain swapping has recently been associated to cystatin C amyloid angiopaty [115] and to Unverricht-Lundborg disease [120]. Staniforth R.A et al. [121] discussed the way in which the domain swapped dimer of cystatins could propagate into fibrillar structure. Human cystatin A and B are found to form amyloid fibrils *in vitro* too, even if with different propensity [122].
It has been found that some plant cystatins, under certain conditions are able to *in vitro* self-aggregate generating typical amyloid fibrils. The fold of plant cystatin resembles that of a pathologically relevant human cystatin, the human cystatin C. For

these reasons, plant cystatins are thought to be a useful model to study the protein self-aggregation phenomenon and the amyloid fibril formation.
The first report on the aggregation of a plant cystatin in amyloid fibrils concerns monellin [123]. It was found that the isolated B chain, when treated under harsh conditions [124], is able to form *in vitro* amyloid fibrils with the typical tinctorial and morphological properties. The observation of amyloid fibrils formation from a plant protein strongly support the view that amyloid-like aggregation is a universal phenomenon occurring in a very wide range of protein species [125, 126].
Cystatins from plants, homologous to animal cystatins, are referred to as phytocystatins [127]. Oryzacystatin (*oryc*) is the first well defined phytocystatin [128, 129]. It occurs in the endosperm of the rice *Oryza sativa L. japonica*, where it shows a potent inhibitory activity towards papain (EC 3.4.22.2) and other cysteine proteinases. Oryzacystatin is similar to type I animal cystatins, in that it consists of 102 amino acids and lacks disulfide bonds, even if its tertiary structure is more similar to those of the type 2 animal cystatins.
Another well characterized plant cystatin is monellin, isolated from the fruit of the tropical West African berry *Dioscoreophyllum cumminsii* [130, 131]. It is a sweet-tasting protein composed of two subunits, one being 45 residues (A chain) and the other 50 residues (B chain) [132]. X-ray and NMR analyses have resolved the structure of monellin showing the typical cystatin superfamily fold with five anti-parallel β-sheets and one α-helix [133, 134, 135]. However, monellin structure lacks of the N-terminal region of the cystatins that is known to host one of the essential molecular determinants for cysteine protease inhibition. As a consequence, monellin lacks of inhibitory activity. The protein is very stable upon acidification, although it is rather easily denatured by temperature or chemical reagents.
In the last years, native monellin has been engineered by fusing the two chains of the protein one to each other. Two mutants were prepared: a single chain monellin (named *SCM*) obtained by joining the C-terminal residue of the B chain directly to the N-terminal residue of the A chain [136], and a single chain monellin (named *mnei*) obtained by linking B and A chains *via* the Gly-Phe dipeptide [137]. These proteins retain both the structure and the sweet taste of native monellin, but their thermal stability greatly increased. Only one of these variants, the SCM, has been analyzed, so far, for its propensity to generate fibrils. It was found that this variant is not able to form amyloid fibrils, in contrast to the native protein.

5. Amyloid and Biotechnology

The amyloid fold, consisting of a continuous β-sheet with β-strands oriented perpendicular to the fibril long axis, has been successfully used for biotechnological applications. Actually, it presents four main advantages.
First, the amyloid structure has a remarkable regularity, being periodical and well ordered.
Second, it permits to explore the nanoscale, manipulating individual atoms and molecules, as proposed by R. Feynman in 1959 ("*I want to talk about the problem of manipulating and controlling things on a small scale*")[138].
Third, using the natural "own" materials, like proteins and peptides, it presents the biocompatibility trait.
Last, it is based on the molecular self-assembly ("molecular Lego") generating structures with a lower complexity than the supramolecular structures [145].
Molecular self-assembly takes place upon molecular recognition, followed by reversible association. It is mediated by weak, non covalent bonds and by

complementarity in shapes among the individual building blocks. Molecular self-assembly realizes the bottom-up strategy [139] using single molecules as building blocks for supramolecular structure. This strategy replaces the top-down approach, i.e. the scale down grinding of large materials to obtain smaller entities. Molecular self-assembly is emerging as a new route to produce novel materials and to complement other materials [146, 147].
Moreover, the dependence of the amyloid formation from the environmental conditions like pH, ionic strength, presence of co-solvents etc, permits to strictly chose and control the best options.
Amyloidogenic polypeptides, like the amyloid β peptide (Aβ) and yeast prion Sup 35, were successfully used to cast metal nanowires and to build conducting nanowires, respectively [140, 141]. Amyloid fibrils of bioadhesive proteins were discovered in natural adhesives and used to create self assembling biomaterials with novel elastic and adhesive properties [142, 143]. Multifunctional peptide fibrils were used as biomedical materials for tissue regeneration and engineering [144]. Novel biomimetic nanocomposites [151] and copolymers [152] were obtained using the high propensity to generate β-sheet based structures of polypeptides from natural spider silk [148, 150]. Several self-assembling peptides and proteins that form nanotubes, helical ribbons and fibers have recently emerged as biological materials, like biocompatible matrices for tissue engineering [154]. Peptide nanofilaments and nanoropes were constructed as smart material for biosensing and signaling upon a useful functionalization [153]. Finally, organogels based on self-assembly were formed by weak non covalent interactions, thus substituting traditional covalent polymers [155].
We planned the production of biologically active fibrils, named catalytic amyloid fibrils. The idea is that of obtaining amyloid fibrils able to transform a substrate in its corresponding product.
This can be achieved by genetically fusing the fibrillogenic moiety to an active enzyme. Two conditions must be realized: (i) the ability of the fibrillogenic partner to form amyloid fibrils even in the presence of the enzyme; (ii) and the ability of the enzymatic moiety in maintaining its activity in the amyloid-like aggregate. Considering the ordinate and polymeric structure of the amyloid, the catalytic amyloid fibrils could become a new type of noncovalent catalytic matrix.
Few reports have been published on fusion proteins containing fibrillogenic moieties.
Serio et al. fused the green fluorescence protein (GFP) to the fibrillogenic domain of the yeast Sup 35 protein and found fibrillar structures associated to green fluorescence [157]. Baxa et al. fused four different enzymes to the Ure2p amyloidogenic protein from *S. Cerevisiae*, with the aim to explore the role of the protein context on the fibrillogenic behaviour of Ure2p. They found that, while the Ure2p fibrillogenic behaviour varied when linked to different enzymes, in amyloid-like filaments the activity of the appended enzyme was found to be reduced [156]. Finally, designed amyloid-like fibrils generated by the well-characterized enzyme RNase A were found to contain native-like molecules capable of enzymatic activity [158].
This opens the way to the construction of new amyloid-based matrices with unique mechanical and biological properties, having the amyloid fibrils a mechanical strength and resistance comparable to those of the steel [159], and being protease and detergent resistant.
Recent studies demonstrated that amyloid fibrils of bacteria, fungi, insects, invertebrates and humans are associated to various activity (functional amyloid), [49, 61], being able to promote different reactions without any associated toxicity.

AIM OF STUDY

The aims of my research activity were the following:

- construction, isolation and characterisation of a recombinant form of the fibrillogenic domain of ApoA-I, a natively unfolded polypeptide corresponding to residues 1-93, responsible for fibril formation *in vivo*;
- definition of the fibrillogenic potential of the ApoA-I domain by conformational analyses;
- construction, isolation and characterisation of 8 recombinant mutated versions of the fibrillogenic polypeptide, associated to amyloidosis;
- analyses on conformational dinamics and stability of plant cystatins, as model proteins to study the fibrillogenic process of human cystatins associated to amyloidosis;
- use of fibrillar polymers to generate catalytic fibrils for biotechnological applications.

METHODS

1. Molecular cloning

Restriction endonucleases were purchased from New England Biolabs (Beverly, MA). Reagents for molecular cloning were from Promega (Madison, WI).

2. Bacterial Strains

Escherichia coli strains BL21(DE3) and JM 109 (purchased from EMD Biosciences, Inc., Novagen Brand, Madison, WI) were used.

3. Vectors

The pGEX-4T-3 expression vector (Amersham Biosciences, Uppsala, Sweden) (Fig. 7), was used to transform *E. coli* strains [160]. The pGEX plasmids are designed to allow inducible, high level intracellular expression of genes, or gene fragments, as fusions with *Schistosoma japonicum* glutathione S-transferase (GST). pGEX-4T-3 expression vector contains a *tac* promoter for chemical induction; a multiple cloning site (MCS); an internal *lac Iq* gene compatible to any *E. coli* host; a thrombin protease recognition site for cleaving the fusion protein.

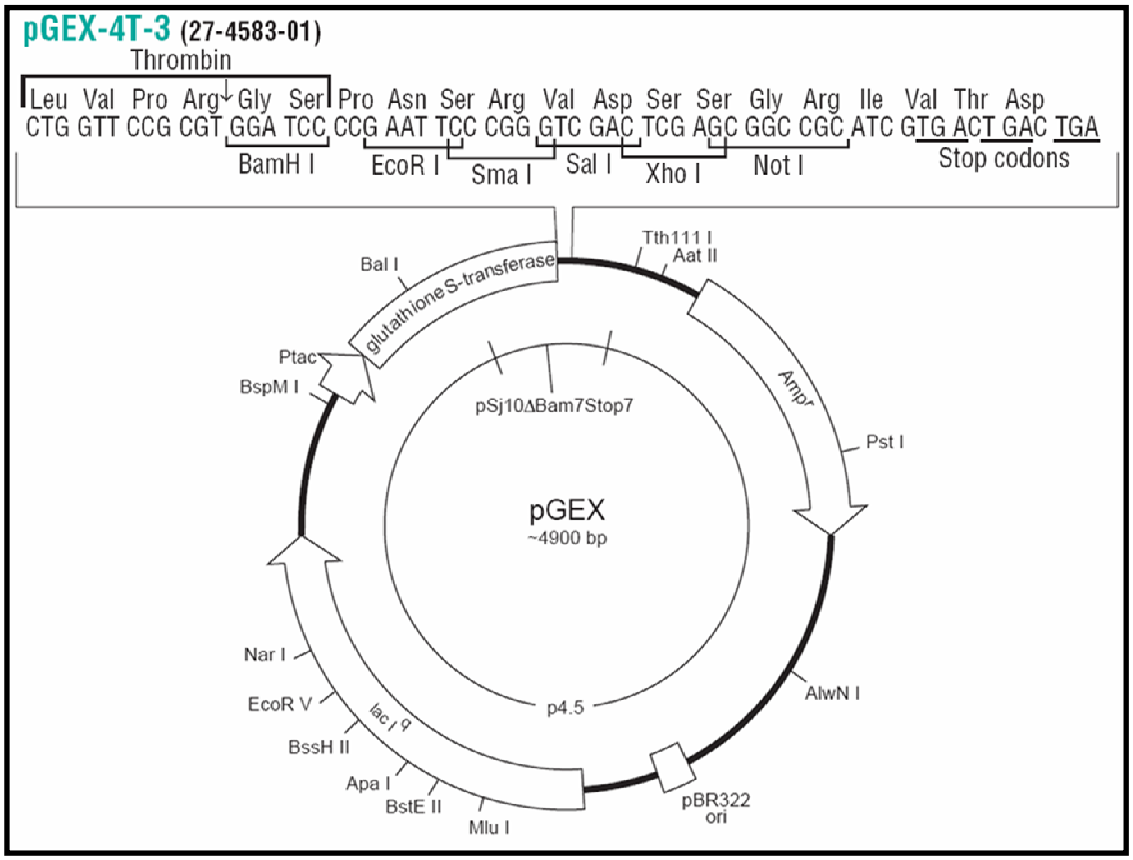


Fig. 7: Structural organisation of pGEX-4T-3 expression vector (Amersham Biosciences, Uppsala, Sweden)

4. Antibiotics

Ampicillin (100 mg/ml) was solubilized in deionized water, filter sterilized and stored at –20°C until use. Ampicillin was used at 100 µg/ml in both solid and liquid media.

5. Solid and liquid media for bacterial strains

LB (Luria-Bertani) and 2x YT-G media were prepared as described by Sambrook and Russel [161]. To prepare a solid medium, 1.2–1.5 g of agar were added to 100 ml medium. Culture media were supplemented with antibiotic. Ampicillin (100 mg/ml) was added to liquid or solid 2x YT-G medium to a final concentration of 100 µg/ml.

6. Antibodies

The antibodies used in the current study were the following: anti-human ApoA-I polyclonal antibodies (Dako-Cytomation, Glostrup, Denmark, dilution 1:500); Anti-GST polyclonal antibodies (General Electric Company, Fairfield, CT, dilution 1:1000), as primary antibodies. Horseradish peroxidase-conjugated goat anti-rabbit immunoglobulin antibodies (Pierce, Rockford, IL, dilution 1:10000); alkaline phosphatase-conjugated rabbit anti-goat IgG immunoglobulin antibodies (General Electric Company, Fairfield, dilution 1:3000), as secondary antibodies. Antibodies were diluted in PBS buffer, containing 0.1% Tween 20 and 1% BSA.

7. Protease inhibitors

Complete Protease Inhibitor Cocktail Tablet (Roche) was used as a mixture of protease inhibitors, according to manufacturer's instruction.

8. Preparation of Bacterial Competent Cells and Transformation

Single clones of BL21 (DE3) *E. coli* strain, grown at 37ºC in LB agar, were inoculated into 5 ml of LB medium. Cells were incubated at 37ºC on a shaker, and grown up to $O.D._{600}$ = 0.6. Cells (1.5 ml) were then incubated 10 min at 0ºC, and collected by centrifugation at 5000 rpm for 5 min at 4ºC. Cell pellet was dissolved in 750 µl of ice-cold 50 mM $CaCl_2$, and incubated for 20 min at 0ºC. Cells were then centrifuged, and the cell pellet was dissolved in 150 µl of 50 mM $CaCl_2$. Competent cells (150 µl) were transformed with plasmidic DNA and incubated 40 min at 4ºC. Heat shock was performed by incubating cells 2 min at 42ºC followed by 2 min at 4ºC. Following the addition of 1 ml of medium, cells were incubated 1 hour at 37ºC and centrifuged at 5000 rpm for 3 min at room temperature. Medium (1 ml) was then removed and the cell pellet was dissolved in the remaining solution (about 100 µl) and plated on LB agar containing ampicillin (final concentration 100 µg/ml).

9. Construction of the chimeric cDNA encoding [1-93]ApoA-I and its variants.

The cDNA encoding fragment 1-93 of ApoA-I (denoted as [1-93]ApoA-I) was obtained by PCR amplification using the full length ApoA-I cDNA as a template, the forward primer a (denoted as F, a in Table III), and the reverse primer b (R, b in Table III). EcoRI and BamHI restriction sites are underlined.

PCR reactions were performed in a final volume of 50 µl, using 20 ng of template DNA. The reaction mixture contained the 2 primers (20 µM each), dNTPs (0.2 mM each), $MgSO_4$ (1 mM), the amplification buffer and Pfx enzyme from *Pyrococcus* (5 U) (Invitrogen Life Technologies, Inc., Carlsbad, CA).

PCR was performed using an Applied Biosystems (Foster City, CA) apparatus, following the procedure indicated below:

Denaturation step: 2 min at 95º

Annealing: 2 min at 55ºC
Elongation: 2 min at 73ºC
for 30 cycles. The cDNA encoding fragment 1-93 of ApoA-I, carrying the mutation L90P([1-93]ApoA-I) was obtained by PCR amplification performed under the same conditions described above, using full length ApoA-I cDNA as a template, oligonucleotide a as the forward primer (Table III) and the mutagenic oligonucleotide, reported in Table III as the reverse primer. Restriction sites are underlined; the P90 codon is in bold.

Mutation	Mutagenic oligonucleotides	Tm (°C)
Wild type	(F, a) 5'CGCGGATCCGATGAACCCCCCCAGAG3'	56
	(R, b) 5'CCGGAATTCTTACACCTCCTCCAGATCCTTG3'	56
G26R	(R) 5'GTCTCT**GCG**GCTGTCTTTGAG3'	62
	(F) 5'GACAGC**CGC**AGAGACTATGTG3'	62
T50R	(F) 5'GACAA**CCG**GGACAGCGTGACC3'	66
	(R) 5'GCTGTCC**CGG**TTGTCAAGGAG3'	64
L60R	(F) 5'AGCAAG**CGG**CGCGAACAGCTC3'	66
	(R) 5'TTCGCG**CCG**CTTGCTGAAGGT3'	64
L64P	(F) 5'GAACAG**CCC**GGCCCTGTGACC3'	68
	(R) 5'AGGGCC**GGG**CTGTTCGCGCAG3'	70
L75P	(F) 5'GATAAC**CCG**GAAAAGGAGACA3'	58
	(R) 5'CTTTTC**CGG**GTTATCCCAGAA3'	58
Δ60-71/ insValThr	(F) 5'CTCCACCTTCAGCAAG**GTCACC**TGGGATAACCTGGAAAAG3'	75
	(R) 5'CTTTTCCAGGTTATCCCA**GGTGAC**CTTGCTGAAGGTGGA3'	75
Δ70-72	(F) 5'GGCCCTGTGACCCAGGATAACCTGGAAAAGGAG3'	73
	(R) 5'CTCCTTTTCCAGGTTATCCTGGGTCACAGGGCC3'	73
L90P	(R) 5'CCGGAATTCTTACACCTCCT**CCG**GATCCTTGCT3'	72

Table III: oligonucleotides used as primers in the PCR reactions performed to generate the [1-93]ApoA-I variants. Melting temperature (Tm) values calculated for each oligonucleotide are shown. (R) and (F), reverse and forward primers, respectively. Restriction sites are underlined; the mutated nucleotide sequences are in bold.

The cDNA encoding the other 7 variants of [1-93]ApoA-I, that have been expressed, reported in Table III, were obtained following a 2-step site-directed mutagenesis, a procedure reported in Fig. 8. In the first step, two PCR reactions were performed in parallel experiments using the following oligonucleotides; d (see Fig. 8) was used as the forward mutagenic primer in combination with oligonucleotide b as the reverse primer, or oligonucleotide c as the reverse mutagenic primer in combination with oliogonucleotide a as the forward primer. In the second step, equivalent amount of

the obtained products were mixed and PCR amplified using primers a and b. Amplification reactions were performed as previously described. The sequence of the mutagenic primers is reported in Table III. All amplification products were analysed by 1% agarose (Promega) gel electrophoresis.

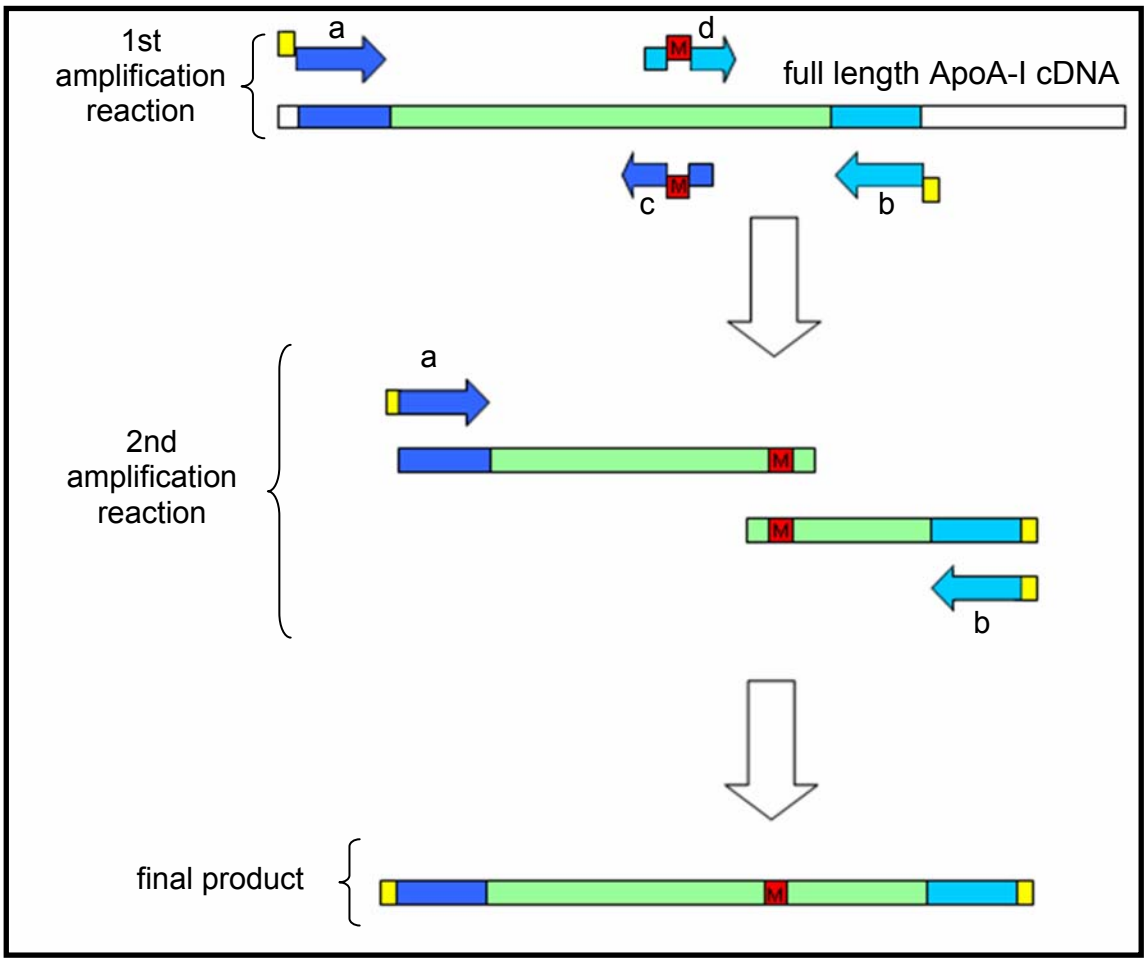


Fig. 8: Schematic representation of the 2-step site-directed PCR mutagenesis for the production of [1-93]ApoA-I variants. Restriction sites are in yellow; the oligonucleotides c and d harbour the mutation (M, red square)

PCR products were purified using a DNA purification system (Promega Biosciences Inc., San Luis Obispo, CA), and digested with EcoRI and BamHI restriction enzymes. Each amplified fragment (6 μg) was digest with 20 U of restriction enzymes for 2 hours at 37ºC. Following the digestion, each fragment was cloned into the corresponding sites of the pGEX-4T-3 expression vector, downstream to the sequence encoding glutathione S-transferase (GST). To this purpose, the expression vector was previously digested with EcoRI and BamHI restriction enzymes, and treated with shrimp alkaline phosphatase (SAP, 20 U) (Roche Applied Science, Indianapolis, IN) for 15 min at 37ºC. SAP enzyme was then inactivated at 65ºC for 10 min. Ligation was then performed using the commercial kit purchased by Promega

Biosciences Inc. (San Luis Obispo, CA), with a 1:10 molar ratio (vector:insert DNA), 1 mM ATP and T4 DNA ligase (1 U) in a final volume of 10 μl.
Following the ligation, the recombinant products were fully sequenced by MWG Biotech, Ebersberg, Germany. Sequence analyses confirmed the expected DNA sequences.

10. Expression of the fusion protein GST-[1-93]ApoA-I and its variants.

Single clones of *E. coli* BL21DE3 strain, previously transformed with one of the recombinant pGEX-4T-3 expression vectors and grown at 37ºC on LB agar containing 100 μg/ml ampicillin, were inoculated in 100 ml of 2x YT-G medium, containing the same antibiotic and 2% glucose. Cells were grown at 37ºC on a shaker up to $O.D._{600}$ = 0.6, diluted with 10 volumes of medium, inoculated in the same medium containing the antibiotic, and grown into a BioFlo 3000 bench top fermentor (New Brunswick Scientific, NY). During cell growth, pH was kept constant to 7.0, temperature at 37°C and shaking within the range 200-300 rpm. DOT was continuously monitored and the foam formation was controlled by adding the antifoam 204 (Sigma Aldrich co.). All operations were conducted aseptically. Following bacterial growth up to $O.D._{600}$ = 1, the expression of chimeric proteins was induced by the addition 0.1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) for 2 h at 37°C. To verify the chimeric protein expression, an aliquot of cell culture was removed before and after induction and analysed by polyacrylamide gel electrophoresis and by Western blotting.

11. Purification of recombinant [1-93]ApoA-I polypeptide and its variants.

Bacterial cells were harvested by centrifugation at 6000 rpm for 20 min at 4ºC. Cells were resuspended on ice-cold PBS, containing protease inhibitors to avoid peptide degradation. Bacterial lysates, obtained by sonication with a Misonix (Farmingdale, NY) apparatus, were incubated for 30 min at 4°C in the same buffer containing 1% Triton X-100. Bacterial lysates were then centrifuged and the supernatant (soluble fraction) was fractionated by affinity chromatography on a glutathione-agarose column (GSTrap, GE) equilibrated in PBS, using the AKTA Purifier chromatographic system (Amersham, Biosciences). GST-containing species were eluted with 50 mM Tris-HCl at pH 8.0, containing 10 mM glutathione following the manufacturer's protocol, but in the absence of DTT to avoid inhibition of thrombin enzymatic activity. Following the chromatography, the column was routinely washed with PBS solution containing 1 mM DTT. The affinity selected proteins were then digested with 8 U/mg of thrombin (Sigma, St. Louis, MO) for 30 h at 4°C in the chromatography elution buffer.
The proteolytic products were separated by HPLC reverse chromatography on a C_8 column (Ultrapure C_8, Beckman Coulter, Fullerton, CA) with a gradient of acetonitrile in the presence of 0.1 % TFA (buffer B) at a flow rate of 1 ml/min, using the Perkin Elmer Series 200 chromatographic system (Perkin Elmer). For each variant of [1–93]ApoA-I, optimal elution conditions were set up by varying the elution parameters. To avoid precipitation and/or aggregation of the polypeptides, recombinant 1-93 polypeptides were neutralized by the addition of 2% ammonium hydroxide and lyophilized. [1–93]ApoA-I polypeptides were analysed by SDS–PAGE on 15% acrylamide gels. For Western blot analyses, anti-human ApoA-I polyclonal antibodies and an anti-GST monoclonal antibodies, coupled to a chemiluminescence detection system (West Pico, Pierce), were used. Purity and identity of [1–93]ApoA-I polypeptides were assessed by mass spectrometric analyses. The polypeptides were

then lyophilized and stored at -70°C until use. For experimental purposes, polypeptides were dissolved in phosphate buffer at pH 7.4 and centrifuged before use.

12. Gel electrophoresis under denaturing conditions

Gel electrophoresis under denaturing condition (SDS-PAGE) was performed as described by Laemmli (1970) [162] on 15% acrylamide gels. Following gel electrophoresis, gels were incubated for 30 min at room temperature in 20% isopropanol, containing 10% acetic acid and 0.1% Coomassie Brilliant Blue (Sigma, St. Luois, Missouri). Gels were then treated with 20% ethanol, containing 7% acetic acid. This solution was replaced by ultra pure water when protein bands were detected.

13. Western blot analyses

For Western blot analyses, following gel electrophoresis, proteins were transferred overnight at 4°C to a polyvinylidene fluoride (PVDF) membrane (Immobilon-P, Millipore, Billarice, MA) using a 25 V electric field. Membranes were incubated in the blocking solution (5% BSA in PBS buffer, containing 0.1% Tween 20) at room temperature for 1 hour. Membranes were then washed with PBS buffer containing 0.1% Tween 20, and incubated either with anti-human ApoA-I polyclonal antibodies, dilution (1:500) or with the anti-GST monoclonal antibodies (dilution 1:1000) at room temperature for 1 h. Membranes were then washed with buffer and incubated on a shaker for 1 hour at room temperature, either with horseradish peroxidase-conjugated goat anti-rabbit immunoglobulin antibodies (dilution 1:10000), or with alkaline phosphatase-conjugated rabbit anti-goat IgG immunoglobulin antibodies (dilution 1:3000) as the secondary antibodies of anti-ApoA-I or anti-GST antibodies, respectively. Membranes were then washed with buffer. The detection of immunopositive species by enzyme-linked chemiluminescence (enhanced chemiluminescence: ECL) was performed according to the manufacturer's instructions (Super Signal®West-Pico Chemiluminescent Substrate, Pierce, Rockford, IL), using a Phosphoimager (Biorad).

14. Isolation of the fusion protein GST-[1-93]ApoA-I as a pure product to generate catalytic fibrils

Following the affinity chromatography previously described, to purify to homogeneity the fusion protein GST-[1-93]ApoA-I, ion exchange chromatography was performed. The GST-containing species eluted from GSTrap were dialyzed against buffer A (3 mM sodium phosphate, sodium acetate and glycine-HCl, pH8). Anionic exchange chromatography was carried out using an AKTA Purifier chromatographic system (GSTrap, GE), on a Resource Q column equilibrated in buffer A. The elution buffer was NaCl 0.5 M (buffer B). Protein species were eluted using the following procedure: a gradient from 0 to 37% buffer B (15 column volumes); an isocratic elution at 37% buffer B (10 vol.); a gradient from 37% to 55% buffer B (35 vol.); a gradient from 55% to 100% buffer B (5 vol.). SDS PAGE ad Western blotting analyses, using either anti-human ApoA-I antibodies anti-GST antibodies (General Electric, CT), were carried out as previously described.

15. CDNB assay

Activity of GST-[1-93]ApoA-I fusion protein was tested using the spectroscopic CDNB (1-chloro-2,4-dinitrobenzene) enzymatic assay. The CDNB reaction mixtures

(500 μl) contained: 50 μL of 1M KH_2PO_4 buffer pH 6.5, 5 μL of 100 mM CDNB, 5 μL of 100 mM glutathione and 440 μL of distilled H_2O. Upon addition of the protein to be tested, absorbance variations were recorded at 340 nm within 5 min and expressed as ΔA_{340}/min. As a small fraction of glutathione reacts with CDNB even in the absence of enzyme, the amount of spontaneously formed product was determined by performing assays in the absence of the enzyme. The obtained ΔA_{340}/min values were subtracted by the corresponding values determined in the presence of the enzyme.

16. Expression and purification of recombinant GST and cystatins

S. japonicum GST protein was expressed using the expression vector pGEX-4T-3 and purified according [160].

Mnei and oryc proteins cDNAs, provided by Dr. A. Pastore of NIMR, MRC, London, were expressed and purified according to Spadaccini et al. [136].

Bradford assay (Pierce) to determine protein concentration was used according to manufacturer's instruction.

17. CD spectra

[1-93]ApoA-I far-UV circular dichroism spectra were obtained in collaboration with the group of Prof V. Bellotti, University of Pavia, using a Jasco J-710 spectropolarimeter (Great Dunmow, Essex, UK), equipped with a temperature control system, using a 1-mm quartz cell in the far UV range 190-250 nm. Each spectrum was the average of three scans with the background of the buffer solution subtracted. Measurements were performed at 20°C at a protein concentration of 0.3 mg/ml solubilized in 3mM glycine, 3 mM sodium acetate and 3 mM sodium phosphate at pH 7 (buffer A). The acidification or neutralization of the protein solution were carried out as described [102]. CD data were expressed as mean residue ellipticity (θ).

The CD spectrum of GST and of the GST-[1–93]ApoA-I were carried out in a similar way at a protein concentration of 5 μM in buffer A, at 25°C, using a Jasko J815 CD Spectrometer (Jasko corp.)

Cystatins far-UV circular dichroism spectra were acquired using a cell of 1 mm path-length and a Jasco J-715 spectropolarimeter (Great Dunmow, Essex, UK) equipped with a PTC-348WI Peltier temperature controller. The protein under test was diluted to a final concentration of 0.2 mg/ml either in 50 mM phosphate buffer, pH 7.0, containing 0.1 M sodium fluoride, or in 15 mM glycine buffer, pH 2.5 containing 0.15 M sodium cloride. Far-UV CD spectra were obtained in the range 190-260 nm as the average of at least 16 scans and blank-subtracted. Spectra processing were obtained by using the Spectra Manager software (Jasko corp.) and the SpecPro program (kindly provided by dott. SR Martin from NIMR, MRC, London). The analysis of the secondary structure content was performed following the programs Continll, Selcon3 and CDSSTR of the CDPro program pack (http://lamar.colostate.edu/~sreeram/CDPro/) [163].

18. Thermal unfolding analyses

Cystatins thermal unfolding was performed at 0,2 mg/ml protein concentration either in 50 mM phosphate buffer pH 7.0, containing 0.1 M sodium fluoride, or in 15 mM glycine buffer, at pH 2.5, containing 0.15 M sodium chloride The temperature was increased from 20°C to 90°C at a rate of 2°C/min. The CD signals were acquired at 1°C intervals at the indicated wavelength. The sample temperature was measured using an immersed thermocouple. Thermodinamic unfolding analyses were

conducted using the Curfit program (kindly provided by dott. SR Martin from NIMR, MRC, London). Thermal denaturation data were fitted to the two-state model:
$Y = \{I_N + S_N T\}F_N + \{I_U + S_U T\}F_U$
where T is the temperature (°C), I_N and I_U represent the intercepts and S_N and S_U represent the slopes of the pre- (N) and post- (U) transition baselines, respectively. F_N and F_U are the protein fractions of the total population in the native and denatured state, respectively.
Free energy values used to calculate the fractional populations were obtained from the modified Gibbs–Helmholtz equation [164].
$\Delta G_T = \Delta H_m (1 - T/T_m) + \Delta C_p [(T - T_m) – T\ln(T / T_m)]$
where T_m and ΔH_m are the midpoint melting temperature and the enthalpy at Tm, respectively, and ΔC_p is the difference in heat capacity between the native and denatured states [165]. Values of ΔG_T were calculated using a ΔC_p value fixed at 1.4 kcal/mol K, a reasonable value for proteins of the size of the protein under test [166] because of the difficulty to define ΔC_p by the analysis.

19. Binding of fibrillar structures to 8-anilino-1-naphthalenesulphonate (ANS)

The ANS(350 μM) emission fluorescence spectra in the presence of proteins were carried out at a protein concentration of 3 μM at 25°C in buffer A, using a LS55 luminescence Spectrometer (PerkinElmer). ANS emission fuorescence spectra were recorded in the range of 400–600 nm at the excitation wavelength of 380 nm and the slit widths set at 5 nm. Spectra were recorded before and after acidification to pH 4 followed by neutralization to pH 7 by adding HCl and NaOH, respectively.

20. Thioflavin T assay

[1-93]ApoA-I polypeptide (20 μM) was incubated at room temperature in buffer A at pH 4.0, containing 0.02% sodium azide and 15 μM Thioflavin T (ThT). When necessary, assays in phosphate buffer at pH 6.4 were also performed. ThT fluorescence emission spectra were acquired at 1 h intervals with a scan speed of 300 nm/min, upon excitation at 450 nm. Spectra were recorded in the range 455-600 nm. Excitation and emission slits were set at 5 nm and 10 nm, respectively. The fluorescence intensity values emitted at 482 nm were plotted as a function of time.
ThT fluorescence emission spectra in the presence of GST or of the GST-[1–93] ApoA-I fusion protein were carried out in a similar way at a protein concentration of 10 μM, 25°C in 3 mM glycine, 3 mM sodium acetate, and 3 mM sodium phosphate (buffer A) at pH 6.4.
Cystatins were incubated at 86°C at a concentration of 100 μM in 15 mM glycine buffer, pH 2.5 containing 0.15 M sodium cloride. 7 μL were withdrawn every hour and added to 800 μl of 25 mM phosphate buffer, pH 7.5, containing 20 μM ThT. Immediately after mixing, ThT fluorescence emission spectra were acquired using a SPEX FluoroMax fluorimeter (Jasko corp.) upon excitation at 442 nm, in the range 455-600 nm. Excitation and emission slits were set at 5 mm and 7 mm, respectively.

21. Congo Red assay

Proteins were incubated for about 60 h at room temperature in buffer A or at pH 4.0 or at pH 7.0. The protein was than diluted to 100μg/ml in PBS and the spectrum acquired (spectrum B). Congo Red was then added to a final concentration of 20 μM (14 μg/ml); upon 15 min equilibration at room temperature, the spectrum of Congo Red was acquired (spectrum A). The spectrum of Congo Red (20 μM) in PBS in the

absence of protein was also acquired, as a control (spectrum C). Spectra were acquired in the range 400-600 nm using a Cary 50 Bio UV-visible spectrophotometer (Varian). The ratio of Congo Red (μM) to protein (μg/ml) was about 1:5 according to Klunk WE et al [176]. Differential spectra of the complex Congo Red-protein were obtained by subtracting spectra B and C from spectrum A, using the CaryWinUV software (Varian corp.).

22. Gel-filtration chromatography.

[1-93]ApoA-I (0,2 mg/ml) was incubated for 72 h at room temperature in buffer A at pH 4.0 or pH 7.0. Gel-filtration chromatography was performed on a Superdex G-75 column (GE) equilibrated in buffer A at pH 7, containing 0.15 M sodium chloride.
Cystatins were incubated at 86 °C, or 50 °C, for 0, 3, 6 h at a protein concentration of 100 μM in 15 mM glycine buffer, pH 2.5, containing 0.15 M sodium chloride. Following incubation, gel-filtration chromatography was performed on a Superdex G-75 column (GE) equilibrated in 15 mM sodium phosphate at pH 7, containing 0.15 M sodium chloride.

23. Atomic force microscopy (AFM)

AFM analyses were performed in the laboratory of Prof. A. Gliozzi, University of Genova. AFM analyses of the cystatin mnei, were performed in London. The protein was incubated for 7 h at 85 ºC at 1 mg/ml in a 15 mM glycine buffer pH 2.5 containing 0.15 M NaCl. Solutions were then diluted 20 times and incubated on a freshly cleaved mica surface. Images were acquired using a Nikon eclipsete 2000U microscope and a Nanowizard AFM (JPK instruments) system. An Olympus BL150 cantilever (30 x 50 μm, 0.03 N/m, Olympus, Japan) at a frequency of 8.3 KHz was used.

24. Bioinformatic tools

Protein sequences were processed using the ProtParam (http://www.expasy.ch/tools/protparam.html) software for the calculation of protein instability index [167]. Chou and Fasman algorithm [168] from Prot Scale (http://www.expasy.ch/cgi-bin/protscale.pl) was used to predict the propensity to generate β-sheet secondary structure. Both ProtParam and Prot Scale are available on the ExPASy Server (http://www.expasy.ch) [169]. Tango algorithm [62] from DisEMBL (http://dis.embl.de/) [170] was used to predict the aggregation propensity of proteins and peptides. For Tango (http://tango.crg.es/) analyses, pH values were varied, while temperature was kept constant at 25°C, ionic strength at 0.01M, in the absence of TFE as a cosolvent. Waltz algorithm [171] (http://switpc7.vub.ac.be/cgi-bin/submit.cgi) was also used to predict amyloidogenic regions in protein sequences at neutral or acidic pH values.
Protein structure images were obtained from PDB (http://www.rcsb.org/pdb/home/home.do) using the Pymol program (http://pymol.sourceforge.net/) [172]. Sequence alignments were performed using ClustalW (http://www.ebi.ac.uk/Tools/clustalw/ index.html) [173].

RESULTS

1. Structural and functional studies on the fibrillogenic domain of ApoA-I

1.1 Expression and isolation of recombinant [1–93]ApoA-I as a stable product

The lack of suitable amounts of the amyloidogenìc domain of ApoA-I from natural sources has impaired so far the understanding of the molecular mechanism of ApoA-I related amyloidosis.

Amyloidogenìc ApoA-I polypeptide, corresponding to the 93-residue N-terminal region of ApoA-I, and denoted as [1–93]ApoA-I, is a natively unfolded protein [102]. Attempts to express a recombinant form of the polypeptide, either in prokaryotic or eukaryotic cells, had failed so far, probably due to the rapid intracellular digestion of the unstructured product, and the high instability of the polypeptide. Recently, we succeeded in the production of a recombinant form of [1–93]ApoA-I [174]. The polypeptide was expressed in bacterial cells, following an experimental strategy aimed at reducing the intracellular degradation during its production [174].

The following strategy was used. The polypeptide was expressed as a chimeric protein, obtained by fusing the 93-residue polypeptide to the enzyme glutathione S-transferase (GST).

The cDNA encoding fragment 1–93 of ApoA-I was obtained by PCR amplification using full length ApoA-I cDNA as a template and suitable oligonucleotides to amplify the 1-93 sequence. Restriction sites for EcoRI and BamHI were also inserted to clone the amplified cDNA in pGEX-4T-3 expression vector, downstream to the sequence encoding glutathione S-transferase (GST) [174]. Competent bacterial cells were transformed with pGEX-4T-3 vector carrying the GST-[1–93]ApoA-I construct and induced to express the recombinant protein.

Analyses by SDS–PAGE of bacterial lysates showed the presence of a major protein species in IPTG-induced cells (Fig. 9A, lane 3) with a molecular mass of about 36 kDa, as expected for the chimeric product. This product was specifically recognised by anti-ApoA-I antibodies (Fig. 9B, lane 3). In this sample, additional immunopositive species with a molecular mass lower than that of the chimeric protein were also present (see below).

By affinity chromatography on a GSH-agarose affinity column, the GST-containing species were selected. In Fig. 10, the affinity chromatography elution profile is showed. Analyses by SDS–PAGE (Fig. 9A, lane 4) confirmed that the selected proteins contained exclusively species immunopositive to anti-ApoA-I antibodies (Fig. 9B, lane 4). Coomassie-stained protein bands, corresponding to the immunopositive species, were excised from the gel and digested *in situ* with trypsin. The resulting peptide mixtures were analysed directly by MALDI mass spectrometry in the laboratory of Prof. P. Pucci of the University of Naples. The analyses showed that the main product was indeed the GST-[1–93]ApoA-I chimeric protein (indicated in Fig. 9 by an arrow). In addition, minor species corresponding to C-terminal truncated chimeric products were also identified.

The [1–93]ApoA-I moiety was then released from the chimeric protein by targeted proteolysis, making use of a unique cleavage site for thrombin positioned between the GST and the [1–93]ApoA-I coding sequences. The proteolysis mixture, analyzed

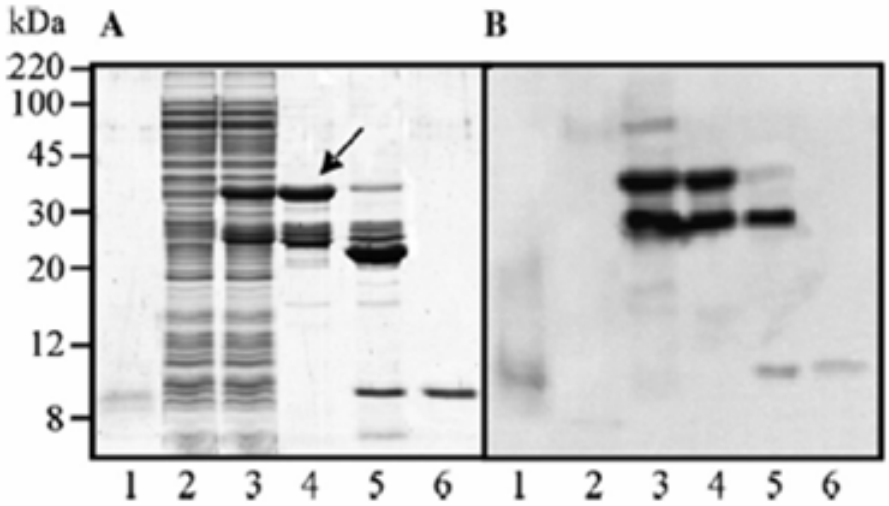


Fig. 9: Analysis by SDS–PAGE of recombinant [1-93]ApoA-I. (A) Coomassie staining; (B) Western blotting.
Lane 1, native [1–93]ApoA-I extracted from ex vivo fibrils;
lanes 2 and 3, soluble fractions prepared from non-induced bacterial cells (lane 2), or from induced cells (lane 3);
lane 4, GST-containing proteins selected by affinity chromatography;
lane 5, thrombin proteolytic products;
lane 6, HPLC purified [1–93]ApoA-I.
The arrow indicates the full length GST-[1-93]ApoA-I fusion protein.

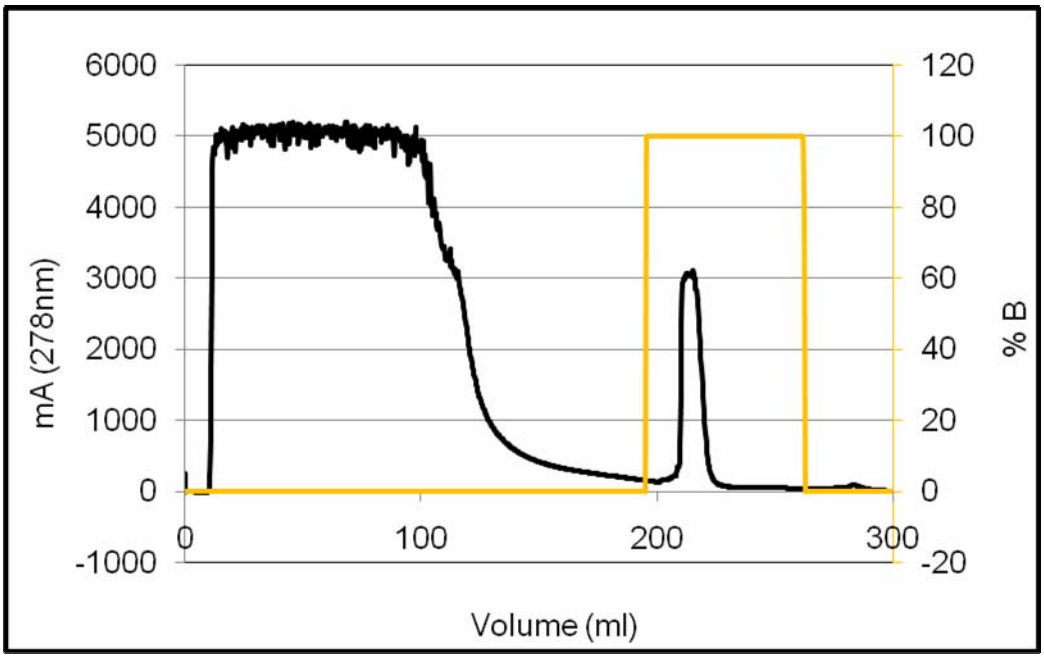


Fig. 10: Affinity chromatography profile of GST-containing proteins selected by a GSTrap column. The recorded absorbance at 280 nm is shown in black; the elution gradient (%B) is shown in orange.

by SDS–PAGE and Western blotting (Fig. 9A and B, lane 5), was resolved by reverse-phase HPLC using a gradient of acetonitrile in TFA, as described in the Methods section.
The eluted fractions were immediately neutralized by the addition of small amount of ammonium hydroxide. The reverse-phase HPLC profile, shown in Fig. 11, showed two well resolved peaks. The minor one (eluted with 48% of acetonitrile and indicated by an arrow) was found to contain a homogeneous protein with the same electrophoretic mobility of the native fibrillogenic polypeptide. This species was recognised by anti-ApoA-I antibodies (Fig. 9A and B, lane 6).

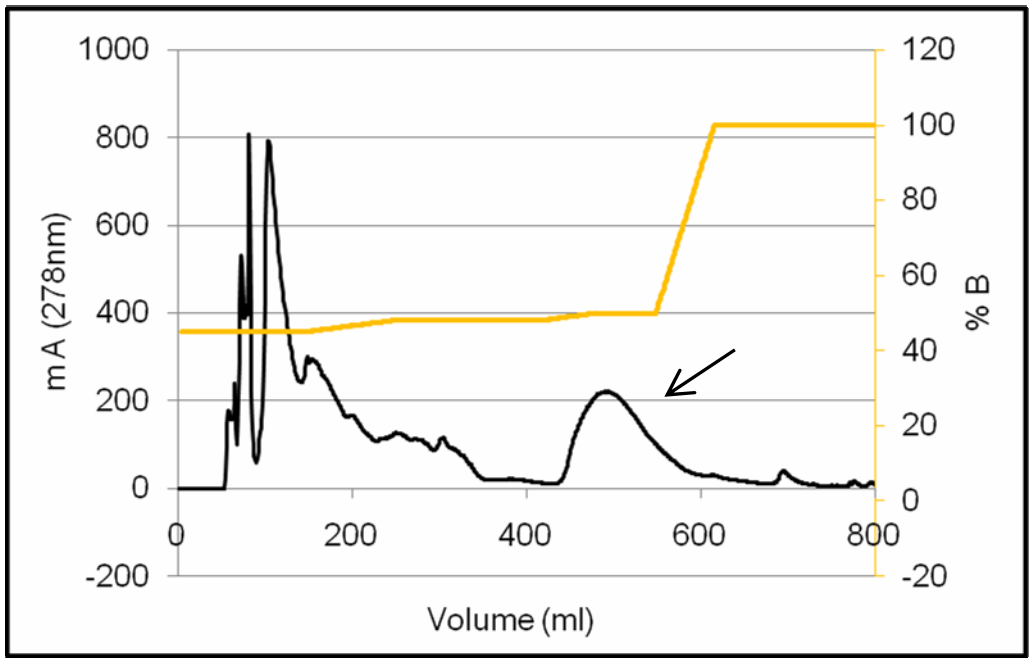


Fig. 11: RP-HPLC profile of [1-93]ApoA-I polypeptide. The absorbance recorded at 280 nm is shown in black; the acetonitrile gradient (%B) is shown in orange. The arrow indicates the peak corresponding to the fibrillogenic polypeptide.

An aliquot of this product was analysed by electrospray mass spectrometry (ESMS). The molecular mass of 10863.8 ± 0.1 Da was calculated and found to perfectly match the expected mass value of [1–93]ApoA-I, with the addition of two extra residues (Gly and Ser). These extra amino acids are in fact present at the N-terminal end of the polypeptide and originate from the chimeric construct following thrombin cleavage.
Using this procedure we were able to isolate a pure recombinant form of [1–93]ApoA-I with a recovery of about 1mg/L of bacterial culture.
Having established a suitable protocol for cloning, expressing and purifying a recombinant form of the amyloidogenic polypeptide of ApoA-I, the way was open to study the self-association process of the amyloidogenìc polypeptide.

1.2 Characterisation of [1–93]ApoA-I

A partial characterisation of natural [1–93]ApoA-I polypeptide, extracted from fibrils obtained from patients affected by a severe form of cardiac ApoA-I amyloidosis, had shown that the 1-93 N-terminal portion of ApoA-I is intrinsically amyloidogenic in a

physiological environment. The natural 1-93 amyloidogenic fragment was found to be able to aggregate in acidic conditions (pH 4.0) generating amyloid fibrils [102].
In collaboration with the research group of Prof. V. Bellotti of the University of Pavia, we analysed the ability of the recombinant [1–93]ApoA-I polypeptide to generate fibrils as its natural counterpart. The fibril formation propensity of [1–93]ApoA-I polypeptide was then investigated using a multidisciplinary approach.

1.2.1 In silico analysis

The aggregation propensity of [1–93]ApoA-I polypeptide was first examined using algorithms proposed by some authors. The instability index of [1–93]ApoA-I (46.67), calculated using Prot Param [167] allowed us to classify the polypeptide as unstable.
The propensity of [1–93]ApoA-I to generate fibrils was first analysed using the Tango algorithm defined by L.Serrano et al. [62].
Tango analyses of polypeptide were performed at 25°C, at 0.01 M ionic strength in the absence of co-solvents. Tango algorithm predicted a very high propensity of [1–93]ApoA-I to generate β-crossed aggregated structures at pH 7.0 and 4.0.
In Fig. 12, the aggregation value (AGG), that describes the overall protein propensity to aggregate, is reported. The aggregation propensity was found to be higher at pH 4.0 than at pH 7.0, according to the experimental data previously reported for natural [1–93]ApoA-I polypeptide [102]. The propensity to generate β-cross aggregates is localised mainly within the region comprised from residue 13 and residue 25 at both pH values (Fig. 12). No propensity to aggregate in a helical conformation was instead predicted by Tango.

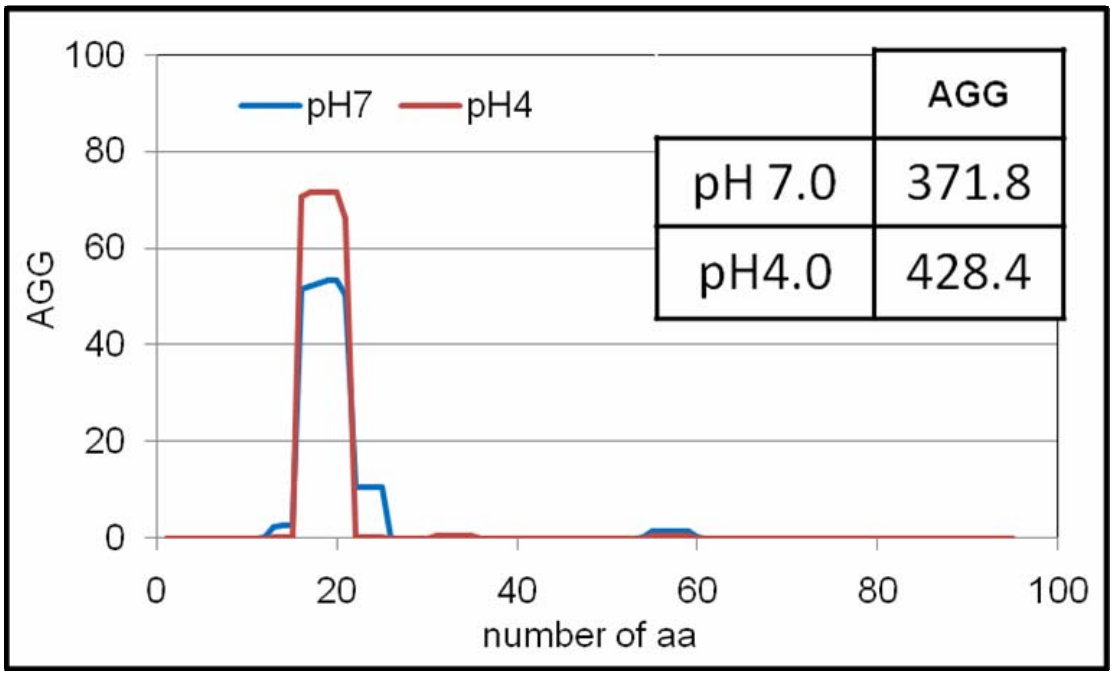


Fig 12: The aggregation propensity of [1-93]ApoA-I residues predicted by the Tango algorithm. The overall aggregation score (parameter AGG), determined by the algorithm for the whole polypeptide at pH 7.0 and 4.0 is shown.

The 13-25 region was predicted to be prone to aggregation also using the Waltz algorithm [171], even if further aggregating regions were also identified, due to the lower specificity of this algorithm (data not shown). Finally, the region 13-25 was predicted to have high tendency to form β-sheet based structures also using the algorithm of Chou and Fasman [168].
Interestingly, the C-terminal region, for which no aggregation propensity was detected using TANGO algorithm, was found to be unaccessible to proteases, both under native conditions and in the presence of a cosolvent (20%TFE), as demonstrated by the group of Prof. P. Pucci of the University of Naples by using limited proteolysis coupled to mass spectrometric analyses [174].

1.2.2 Spectroscopic characterisation of [1-93]ApoA-I

The conformational characterisation of [1–93]ApoA-I was performed by spectroscopic (a) and spectrofluorimetric (b) analyses described below.

a. CD spectroscopic analyses

CD spectroscopic analyses were performed in collaboration with the group of Prof. V. Bellotti, University of Pavia. The overall far-UV CD spectrum of [1-93]ApoA-I at neutral pH is shown in Fig 13A. The minimum at 203 nm indicated that the peptide is highly unstructured under these conditions, which is in line with the *in silico* characterisation. A marked change in the spectrum was observed at pH 4.0, with a shift of the major minimum from 203 nm to 208 nm. This was accompanied by a considerable increase in ellipticity at 222 nm and 190 nm. Such spectral changes are consistent with a transition to a helical structure at pH 4.0, and are in good agreement with the behaviour of the natural [1-93]ApoA-I polypeptide[102]. Moreover, as described for the natural polypeptide, when the protein was left at pH 4.0 we observed the decay of the CD signal during protein aggregation and precipitation.
Interestingly, the pH-induced changes in the spectrum were reversed when the pH was neutralised within a few seconds (Fig. 13A).
Fig. 13B reports the time dependence of the CD spectra changes induced by acidic conditions. After 60 min incubation at pH 4.0, we observed the disappearance of the minimum at 222 nm and the persistence of a second minimum at 205 nm, which is suggestive of a transition to a β-sheet structure.
From these data, a complex and multistep intriguing aggregation pathway emerges for [1–93]ApoA-I. Major conformational rearrangements are involved and different intermediate species with own structural characteristics are generated during the aggregation process. During this pathway, the fibrillogenic polypeptide assumes a random coil structure at neutral pH, shifts into an unstable helical conformation at acidic pH and then aggregates into a β-sheet-based structure.

b. Spectrofluorimetric analyses

To further investigate the transient helical conformer, which represents a key intermediate in the fibrillogenic pathway, we used spectrofluorimetric techniques to analyse the binding of the apolar dye ANS to [1-93]ApoA-I.
The ANS fluorescence dye is frequently used to detect partially folded, molten globule-like intermediate states of globular proteins. This hydrophobic dye binds to

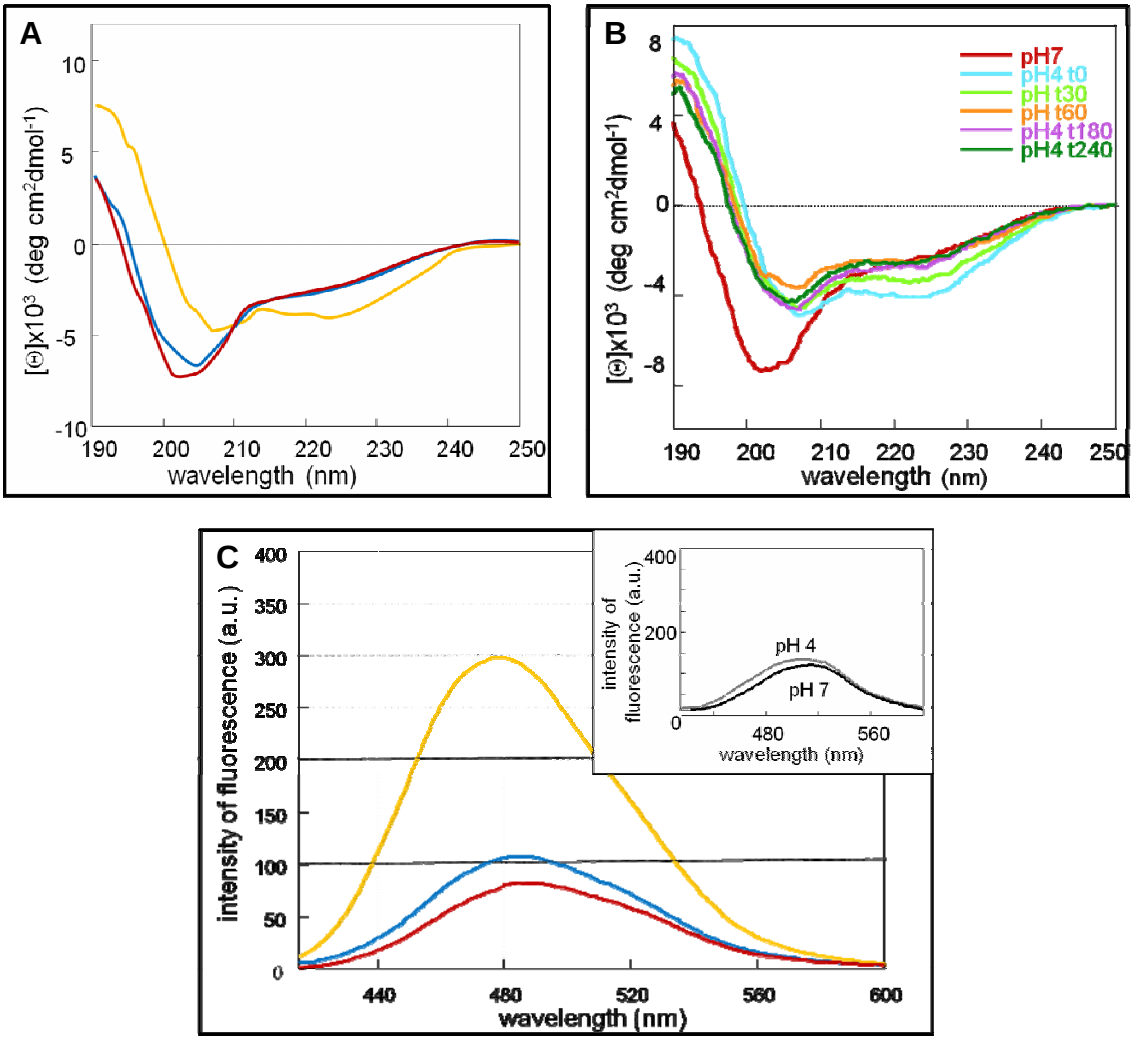


Fig. 13: Spectroscopic analyses of [1–93]ApoA-I. (A) pH-induced transition of [1–93]ApoA-I secondary structure monitored by far-UV CD. Spectra were recorded at pH 7 (blue lines), pH 4 (orange lines), and upon neutralisation to pH 7 (red lines). (B) Time dependence of the induction of secondary structure of [1–93]ApoA-I at pH 4. The far-UV CD spectra were recorded at pH 7 and after a pH shift to pH 4 at the indicated time intervals. After about 60 min of incubation at pH 4, the CD spectra indicated the transition to a β-sheet structure. (C) pH-induced binding of ANS to [1–93]ApoA-I. Symbols are as in (A). ANS emission fluorescence spectra were recorded in the range of 400–600 nm, at the excitation wavelength of 395 nm, with the slit widths set at 5 nm. Spectra of the free dye at pH 7 and 4 are shown in the inset.

solvent-exposed hydrophobic surfaces of polypeptide chains. Folded or unfolded states of proteins exhibit very weak binding affinity for ANS, whereas conformational states with clusters of exposed hydrophobic groups have a much higher affinity, making this dye a suitable marker for partially folded protein states. Upon binding, ANS induces a characteristic enhancement of fluorescence intensity and a blue shift of the emission wavelength [17].

In the case of [1-93]ApoA-I, binding of ANS to the protein at pH 7.0 led to a slight blue shift from 515 to 489 nm, without any increase in fluorescence emission (Fig. 13C). The acidification of the pH, however, led to a substantial change of ANS fluorescence, reflecting the pH-induced transformation into a partially folded conformation. At pH 4.0, the ANS maximum emission shifted to 477 nm with a considerable increase in fluorescence intensity (Fig. 13C), consistent with the behaviour previously ascribed to proteins in a molten globule state [17]. When the pH was returned to pH 7.0, the reversibility of the ANS–ApoA-I polypeptide complex was found to be consistent with the data obtained by the CD spectra. In the inset of Fig. 13C the pH dependence of the fluorescence intensity of the free dye is shown for comparison.

The data obtained by CD and fluorescence experiments have shown that a pH shift from 7.0 to 4.0 induces the formation of a helical conformer, that displays a CD spectrum consistent with a predominant α-helical structure and the ability to bind ANS in a manner similar to that observed for molten globule states. This interaction is abolished when pH was returned to pH 7.0. The conversion of the protein from a random coil structure to a helical/molten globule state at pH 4.0 is a key step in the fibrillogenic pathway. Noteworthy, a helical conformer was identified as an intermediate species in the fibrillogenic pathway of diverse proteins [175]. This intermediate displays a strong propensity to oligomerise. For this reason, attempts to further analyse this intermediate species of the fibrillogenic pathway of some amyloidogenic proteins failed, due to the rapid formation of insoluble species.

1.3 Analyses of [1-93]ApoA-I fibrillar structure

We analysed the presence of a β-sheet-based structure upon long exposure of [1-93]ApoA-I to acidic conditions. To detect the presence of β-cross structure, the typical motif of fibrillar structures, we performed colorimetric assays with Congo Red as well as fluorimetric assays with Thioflavin T. Both assays are commonly used to detect the presence of amyloid fibrils.

1.3.1 Congo Red assays

The ability of amyloid proteins to induce a shift in the spectrum of Congo Red (CR) depends on the protein aggregation state. Upon binding of Congo Red to ordered aggregates, the absorbance maximum of the dye undergoes a red shift from 490 to 540 nm [176]. In addition to this effect, many amyloid fibrils show inherent light scattering properties [177].

We incubated [1-93]ApoA-I at 1 mg/ml in buffer A (3 mM glycine, 3 mM sodium acetate and 3 mM sodium phosphate) at pH 7.0 or pH 4.0. After 3 days, the Congo Red assay was performed.

Spectra of [1–93]ApoA-I (100 μg/ml final concentration) and CR (20 μM) were recorded separately (green and blue lines, respectively, for [1-93]ApoA-I preincubated at pH 4.0, Fig. 14A, or at pH 7.0, Fig 14B). Congo Red (20 μM final concentration) was then added to the protein solution and incubated 15 min to allow the dye to bind to fibrillar structures (Fig. 14A and B, orange lines). The spectrum of the complex Congo Red-protein was then recorded. Differential spectra were obtained by subtracting the spectrum of the dye and that of the protein alone from the spectrum of the complex. The results of these experiments are shown in Fig. 14.

A differential spectrum was obtained when CR was added to the polypeptide previously incubated for 72 h at pH 4.0 (see Fig.14 C, red line). This spectrum

showed the typical red shift described above. No spectral differences were detected when the polypeptide was preincubated at pH 7.0 (see Fig.14 C grey line). This is suggestive of a specific interaction of the dye with ordered β-cross structures, generated during the incubation at pH 4.0.

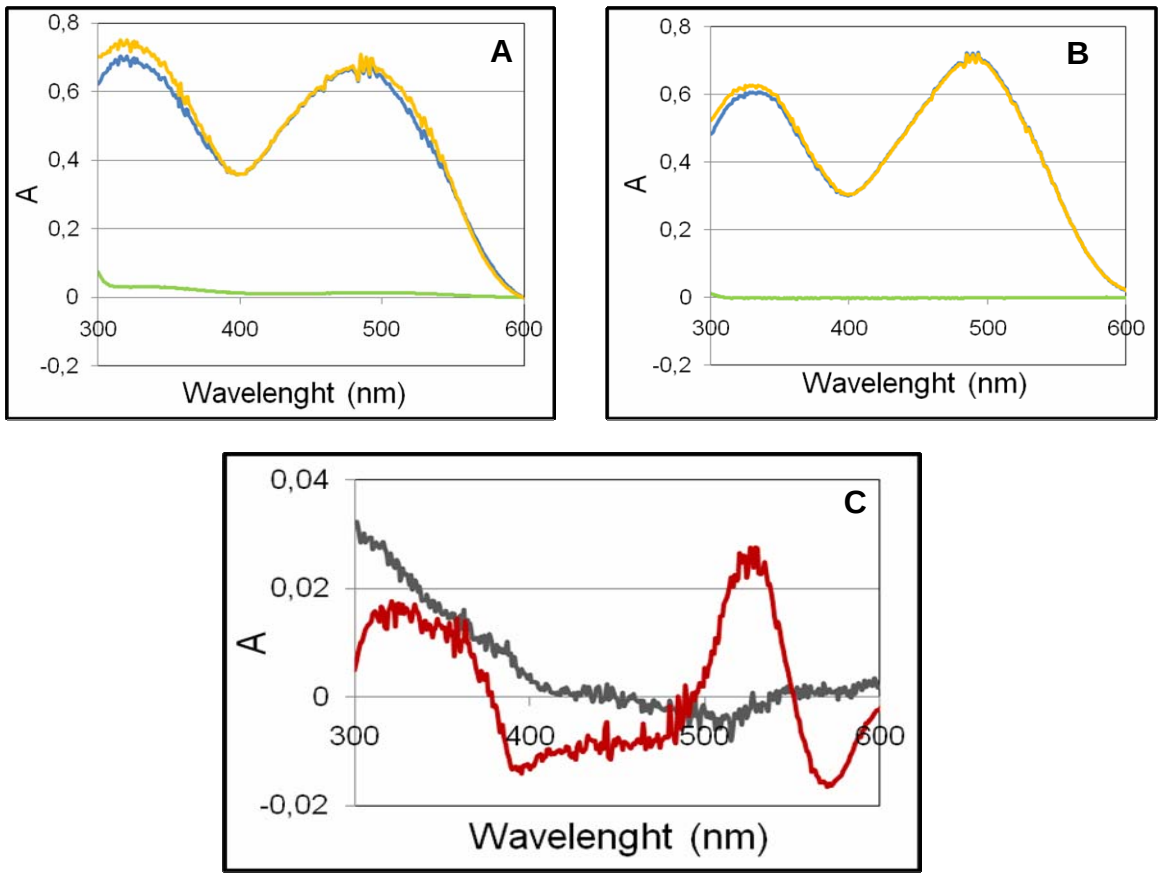


Fig. 14: Congo Red spectra of [1-93]ApoA-I pre-incubated for 72 h at pH 4.0 (A) or at pH 7.0 (B). The spectra of the proteins in the presence of Congo Red are in orange; the spectra obtained with proteins alone are in green; those of Congo Red are in blue. (C) The differential spectra obtained by subtracting the spectra of the protein alone and of Congo Red alone from the spectrum of the protein in the presence of Congo Red at pH 4.0 (red line) or pH 7.0 (gray line).

1.3.2 ThT assays

In spite of a still unknown binding mechanism, Thioflavin T (ThT) is able to bind selectively to amyloid fibrillar structures, and to undergo characteristic spectral changes upon binding to amyloid fibrils [178] These changes consist of a shift from 438 nm to 482 nm of the fluorescence emission major maximum, accompanied by a significant increase of fluorescence up to 80 fold [179].

ThT assays were performed to monitor the formation of fibrils by [1-93]ApoA-I at pH 4.0 and 6.4. The protein (2.5 μM) was incubated up to 300 h in phosphate buffer at pH 6.4, containing 0.02% sodium azide, in the presence of 15 μM ThT. In a parallel experiment, the protein (20 μM) was incubated in 10 mM in glycine buffer at pH 4.0 with ThT 15 μM. ThT fluorescence emission at 482 nm was measured during [1-93]ApoA-I incubation every hour.

Figure 15 shows the time-course of ThT fluorescence emission in the presence of [1-93]ApoA-I. An increase in the ThT fluorescence emission was detected at both pH values, with the characteristic sigmoidal shape suggestive of a cooperative phenomenon. The midpoint of the transition was calculated to be at about 120 h of incubation at pH 4.0, and at about 225 h of incubation at pH 6.4. This indicated that typical polymeric β-cross structure were generated in both conditions, but an acidic environment and high protein concentrations favour the polymer formation.
This is in line with the results published by Prof V. Bellotti [102], who demonstrated that the natural polypeptide is able to form fibrils at pH 4.0, whereas at pH 7.0 fibril formation required much longer times.

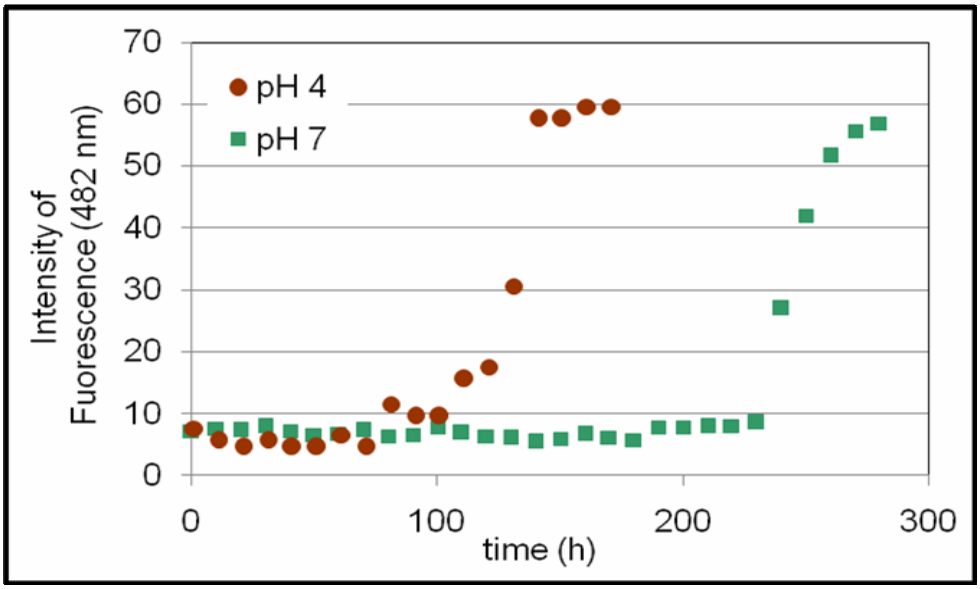


Fig 15: The kinetics of ThT (15 μM) emission fluorescence at 482 nm in the presence of [1-93]ApoA-I. Aggregation was monitored at a protein concentration of 2.5 μM, at 37°C, in 12 mM phosphate buffer pH 6.4 (green squares) or in buffer A at pH 4.0 (red circles).

Taken together, the above reported results revealed that at pH 4 [1-93]ApoA-I converts rapidly from random coil to an unstable intermediate with a helical molten globule state, able to bind ANS. The polypeptide than aggregates into a β-sheet-based polymeric structure, having amyloid-like features and the ability to bind specific dyes such as Congo Red and Thioflavin T.

1.4 Analysis of the aggregation state

Analyses of the aggregated species generated by [1-93]ApoA-I were then carried out by gel-filtration chromatography and AFM.

1.4.1 Analyses by gel-filtration chromatography

The oligomeric state of [1-93]ApoA-I polypeptide was analysed by gel-filtration chromatography. Fig. 16 shows the chromatographic profile of [1-93]ApoA-I, determined for the non incubated protein, and for the protein incubated for 3 days at pH 6.4 (Fig. 16A). The profiles obtained at pH 4.0, are shown in Fig. 16B.

The non incubated [1-93]ApoA-I sample was eluted at an elution volume consistent with a dimeric species (Fig 16A, black line). After three days of incubation at pH 6.4 (Fig. 16A orange line) about 80% of the protein was recovered from the column, as about 20% of insoluble species was generated and removed by centrifugation prior to be analysed. Trace amounts of species with a molecular mass consistent with a monomeric species were also present (arrow in Fig. 16A)
These results indicated that the polypeptide spontaneously associates generating dimers. When [1-93]ApoA-I was analyzed after incubation at pH 4.0 (Fig. 16B, orange line), a lower amount of the dimeric species was observed when compared to the profile obtained with non incubated [1-93]ApoA-I, suggesting that further protein aggregation and/or precipitation occurred, generating high molecular weight species that are not eluted from the column.

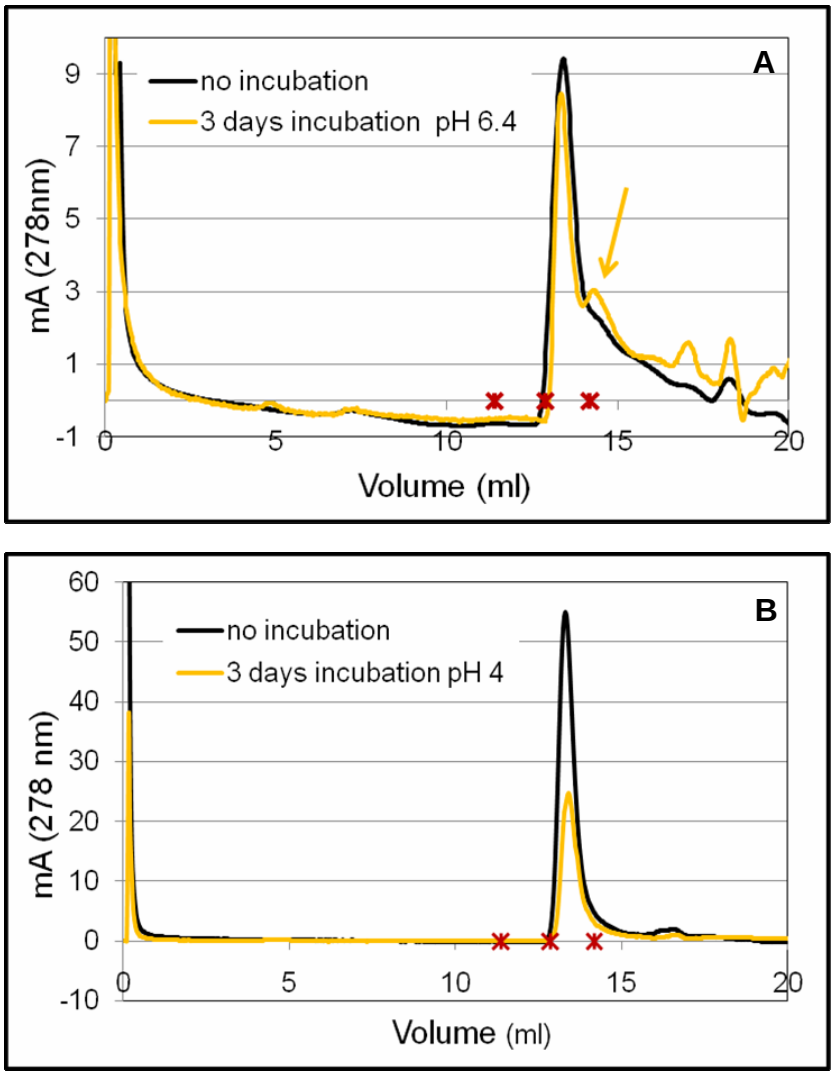


Fig 16: (A) Gel-filtration chromatographic profiles of non incubated (black) or incubated (orange) [1-93]ApoA-I at pH 6.4. The arrow indicates the molecular species eluted as a monomer. (B) Chromatographic profiles of non incubated (black) or incubated (orange) [1-93]ApoA-I at pH 4.0. Asterisks indicate the elution volumes of BSA, α-amylase and cytochrome c used as standards for column calibration.

1.4.2 Analyses by electron microscopy and atomic force microscopy

The data described in this paragraph have been collected by Prof. A. Gliozzi of the University of Genova, which collaborates with our group as an expert of microscopy techniques. I included these data in my thesis in that they are part of the analysis of [1-93] ApoA-I fibrils.

Fibrils generated by [1–93]ApoA-I incubated at pH 4.0 were morphologically analysed. Fig. 17A shows an electron microscopy image obtained upon incubation of recombinant [1–93]ApoA-I under conditions that induce fibrils formation. A nest of narrow protofilaments and well defined fibrils of [1-93]ApoA-I were detectable after 72-90 hours of incubation, consistently with the results obtained with the natural polypeptide [103]. Images obtained by AFM are presented in Fig. 17B. The images clearly indicated that fibrillar structures with the typical amyloid morphology coexist with globular aggregates. The heights of protofilaments and mature fibrils from the AFM images were calculated to be 0.8±0.3 nm and 2.4±0.6 nm, respectively, values similar to those measured for the natural fibrillogenic polypeptide [103].

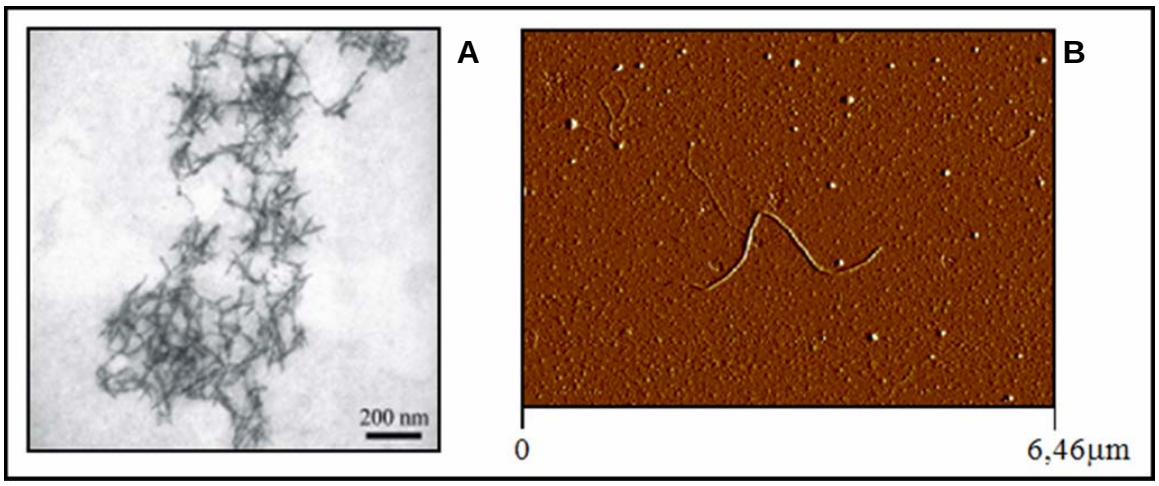


Fig 17: Electron microscopy image (A) and AFM image (B) of [193]Apo A-I fibrils

In conclusions, the data obtained from gel-filtration and microscopy clearly showed that [1–93]ApoA-I has a high propensity to form multimeric aggregates, generating typical amyloid fibrils.

2. [1–93]ApoA-I variants associated to amyloidosis

2.1 Expression and isolation of recombinant [1–93]ApoA-I variants

The experimental strategy to produce the fibrillogenic [1–93]ApoA-I as a recombinant protein can be successfully used to produce the mutated versions of [1–93]ApoA-I. As discussed above, 8 mutations in the N-terminal domain of ApoA-I (called “internal” mutations as they are located within the fibrillogenic domain) have been demonstrated so far to be associated to amyloid pathologies [84]. Following the protocol set up for [1-93]ApoA-I, we planned to express and isolate to homogeneity

the 8 mutant versions of [1-93]ApoA-I polypeptide, thus generating the complete panel of variants of the 93-residue ApoA-I fibrillogenic polypeptides. Six of them harbour a single residue substitution (mutants G26R, T50R, L60R, L64P, L75P and L90P), whereas two mutants harbour a deletion at sequence 70-72 and 60-71, respectively (in the latter case the deleted sequence is replaced by VT) [84]. In order to generate the complete panel of the fibrillogenic polypeptide variants, we mutated, accordingly to the strategy described in details in the Methods section, the cDNA encoding [1-93]ApoA-I. In Table IV (see below) the list of these mutations are shown. Each mutated DNA sequence, encoding the corresponding variant, was obtained by PCR site-directed mutagenesis of the original sequence using suitable primers designed to introduce base substitutions, base deletions or insertions, accordingly.

It has to be underlined that the production of recombinant forms of the polypeptide variants is essential to produce sufficient protein amount to perform studies on the effects of each mutation on the polypeptide stability and propensity to aggregation. No attempts to express these polypeptides as recombinant proteins were previously made.

Variants of [1–93]ApoA-I were expressed in bacterial cells following the same experimental strategy used to produce the [1–93]ApoA-I polypeptide. Polypeptides were all expressed as chimeric proteins, obtained by fusing each 93-residue polypeptide to glutathione S-transferase (GST).

Analyses by SDS–PAGE of bacterial lysates were performed to confirm the expression of the recombinant proteins. In all cases, a major protein species of about 36 kDa, i.e. the molecular mass expected for the chimeric product, specifically recognised by anti-ApoA-I antibodies, was identified (not shown). Each chimeric product was isolated from bacterial lysates using the GST affinity chromatography. Polypeptides were then released by the chimeric construct by targeted proteolysis with thrombin protease. The proteolytic products were resolved by reverse-phase HPLC. Particular attention was paid to choose experimental conditions that avoid precipitation and/or aggregation of the recombinant polypeptides. To isolate each variant, an optimisation of the purification procedure set up for [1–93]ApoA-I has been necessary.

In Fig.18, the HPLC profiles (black lines) obtained for each ApoA-I derived amyloidogenìc variant is shown. The orange lines indicate the elution gradients used to isolate the final products.

Protein fractions were then analyzed by SDS-PAGE and Western blot analyses (Fig. 18A and B, respectively). The results demonstrated that the peak indicated by an arrow in each chromatogram contains a pure protein species with the same electrophoretic mobility of the native fibrillogenic polypeptide (Fig. 18A). These polypeptides were recognized by anti-ApoA-I antibodies (Fig. 18B).

An aliquot of the isolated polypeptides was analysed by electrospray mass spectrometry (ESMS) by the group of Prof. P. Pucci, University of Naples. The molecular mass measured for each recombinant variant is listed in Table IV and found to perfectly match the mass value expected for each [1–93]ApoA-I variant. Furthermore, these analyses indicated than the isolated products were homogeneous (>90% purity).

In Table IV, the amount of pure recombinant products, expressed as mg of recombinant product obtained from 1 L of bacterial culture, is also reported. Using this procedure, we were able to produce recombinant homogeneous forms of the 8 mutants of [1–93]ApoA-I polypeptide.

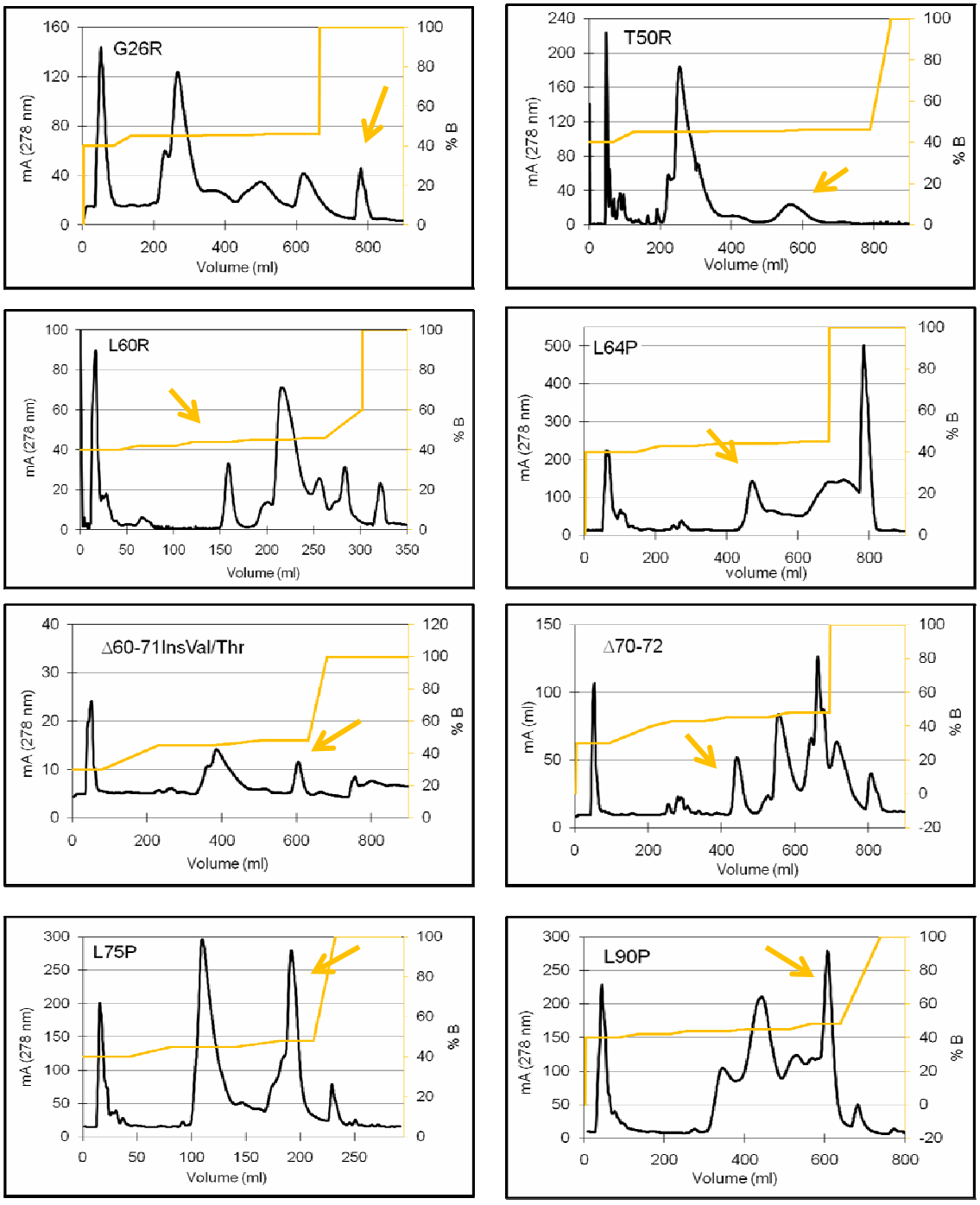


Fig 18: The chromatographic profiles obtained by RP HPLC of the 8 recombinant variants of [1–93]ApoA-I. In orange the elution gradient is shown. The arrows indicate the peaks corresponding to the expected products

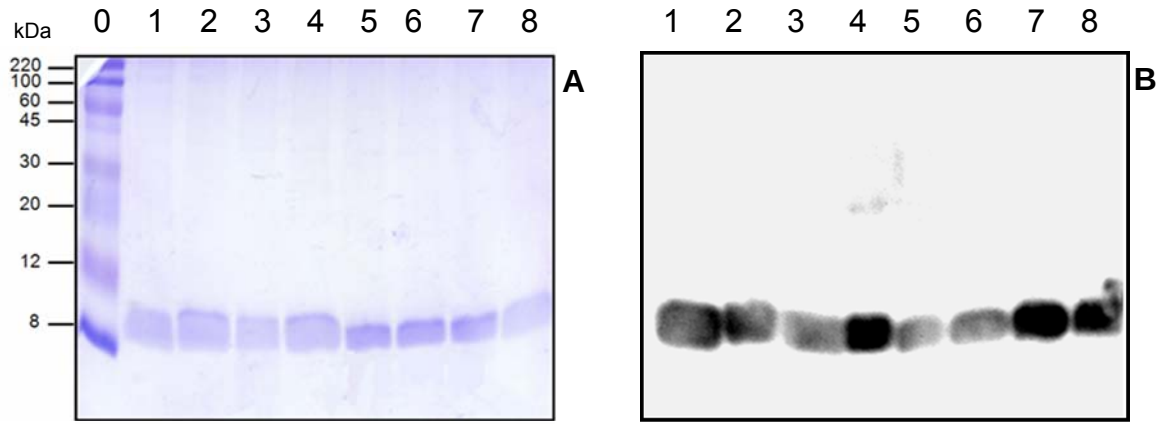


Fig. 19: Analyses by SDS-PAGE (A) and Western Blotting (B) of the proteins isolated by RP-HPLC (Fig. 18)
Lane 0: molecular weight markers; Lane 1: variant G26R; Lane 2: variant T50R; Lane 3: variant L60R; Lane 4: variant L64P; Lane 5: variant L75P; Lane 6: variant L90P; Lane 7: variant Δ70-72; Lane 8: variant Δ60-71 ins Val/Thr

[1–93]ApoA-I variant	Molecular weight	Final product recovery (mg/L of bacterial culture)
G26R	10963.73±0.23Da	0.5
T50R	10834.22±0.26Da	0.3
L60R	10907.16±0.22Da	1.2
L64P	10848.34±0.25Da	0.8
L75P	10847.84±0.45Da	0.8
L90P	10847.64±0.39Da	3.6
Δ70-72	10401.57±0.19Da	0.9
Δ60-71/ins Val-Thr	9665.58±0.22 Da	0.2

Table IV: The 8 mutants of [1–93]ApoA-I are reported with an acronym indicating the mutational event. Their molecular mass and purification yields are indicated.

2.2. Characterisation of [1–93]ApoA-I mutants

Experiments aimed at the characterisation of the conformational state of the 8 amyloidogenic variants of [1–93]ApoA-I, using the techniques previously used to analyse [1–93]ApoA-I fibrillogenesis, are presently in progress. Nevertheless, some preliminary results are available.
This work is aimed at the definition of the impact of single mutations on the protein aggregation propensity. This study would led to define an algorithm able to correlate structural determinants of fibrillogenic polypeptides to their functional features.

Turbidimetry and CD analyses aimed at studying conformational and structural dynamics of mutated polypeptides are presently in progress, in collaboration with the group of Prof. V.Bellotti of University of Pavia. Interestingly, preliminary results, obtained by turbidometric analyses, indicated that almost all mutants display an increased aggregation propensity with respect to [1-93]ApoA-I (Fig. 20). This is in line with the observation that the mutated polypeptides represent the main constituents of amyloid deposits in affected patients. For mutants Δ60-71ins Val-Thr and L64P the highest aggregation propensity was determined.

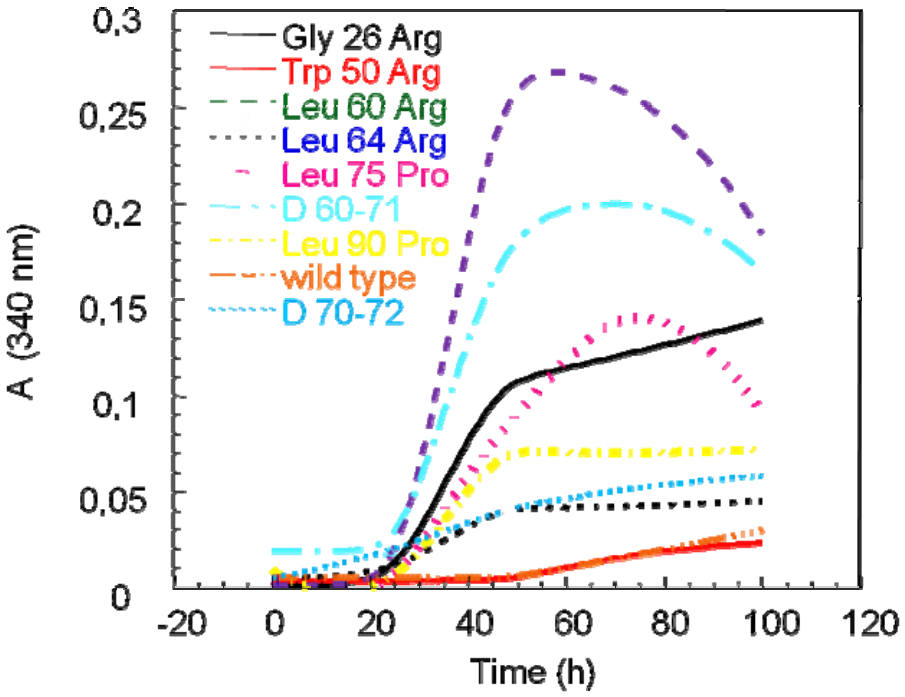


Fig. 20: Turbidimetric assays of the aggregation kinetics of [1–93]ApoA-I recombinant variants.

Moreover, In collaboration with the group of Prof. A. Gliozzi of the University of Genova, AFM analyses of fibrils generated by the recombinant variants are also in progress. Preliminary results suggested that variants generate fibrillar structures, whose morphology will be analysed in details. These analyses are aimed at the definition of the fibrillogenic potential of pathological mutants by microscopic techniques.

We hope that the integration of the results obtained from these independent experimental approaches will be helpful to the definition of the molecular bases of fibrillogenesis.

3. Cystatins and amyloid aggregation

Plant cystatins may undergo the typical amyloid aggregation process [123, 124]. Since they are able to form typical amyloid fibrils and present the conserved structural motif of cystatins, plant cystatins could be an interesting model to shed light on cystatins aggregation.
During my research activity in London, in the Laboratory of Dr. A. Pastore at the Molecular Research Council (MRC), National Institute for Medical Research (NIMR), I studied the aggregation pathway of two plant proteins belonging to the cystatin superfamily: the *mnei* protein, a recombinant single chain monellin engineered from the naturally occurred double chains monellin, from *Dioscoreophyllum cumminsii*, and oryzacystatin I (oryc), a cystatin from *Oryza sativa L. japonica.*

3.1. In silico analyses

Single chain monellin (mnei) and oryzacistatin (oryc) sequences were aligned using Clustal x [173] (Fig. 21A). A comparison of their structure from NMR data (PDB 1FA3 for mnei, [135]: PBB 1EQK for oryc, [181]) is reported in Fig. 21B using Pymol program [172]

Fig. 21: (A) Alignment of mnei and oryc amino acidic sequences. (B) Protein structure superimposition: mnei (blue, PDB 1FA3) and oryc (orange, PDB 1EQK).

As shown in the alignment, the main difference concerns the N-terminal peptide of oryc, involved in the inhibition of cysteine proteases. This segment has no counterpart in mnei protein, accordingly to the finding that mnei lacks protease inhibitory activity.
Moreover, the superimposition of the structures (Fig. 21B) shows further structural differences, i.e. the turns show different length and the C- terminal segments of the

two proteins span in opposite directions. In spite of these differences, the main elements of secondary structure are conserved.
On the basis of the instability index, calculated using Prot Param [167], and found to be 26.63 for mnei and 32.73 for oryc, both proteins were classified as stable proteins. By contrast, when I carried out the analyses of the aggregation propensity of these proteins using Tango algorithm [62], the aggregation propensity of oryc was found to be higher than that of mnei, both at pH 4.0 and 7.0 (data not shown). This is not in contrast with the observation that for natural monellin amyloid fibrils were observed [123, 124] as Tango algorithm predicts aggregation rather than amyloid fibril formation [62].

3.2 Far-UV CD analyses

For CD analyses, proteins were tested at a concentration of 0.2 mg/ml either in 50 mM phosphate buffer, pH 7.0, containing 0.1 M sodium fluoride, or in 15 mM glycin buffer, pH 2.5, containing 0.15 M sodium cloride. Spectra were recorded in the range 190-260 nm.
Far-UV CD spectra of mnei were measured at pH 2.5 and 7.0, i.e. pH values below the pI of the protein (Fig. 22). Spectra analysis indicated that the protein has a high β–sheet content in both conditions, with a negative band at about 215 nm. Thus, no pH-induced conformational transition occurred. Nevertheless, at pH 2.5 the minimum increases in intensity, suggesting that protein aggregation and precipitation occurred, possibly induced by the destabilizing action of acidic pH.

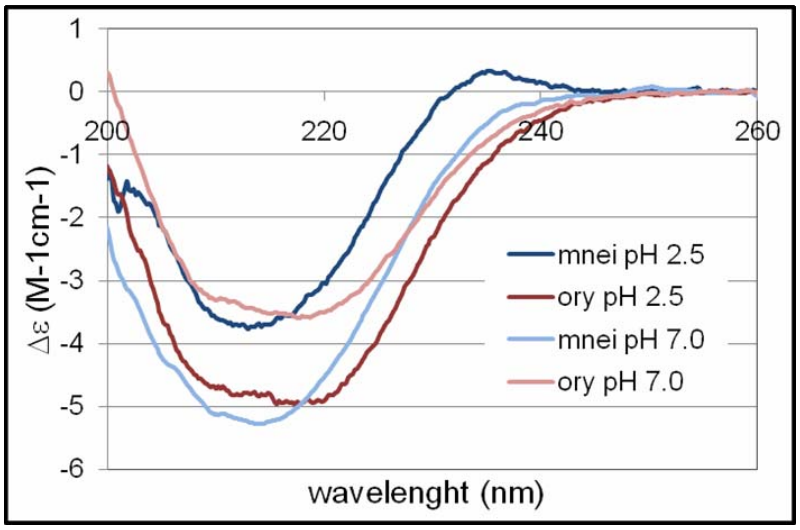


Fig. 22: CD spectra of mnei (blue lines) and oryc (red lines) in phosphate buffer at pH 7.0 (dark colours) or in glycine buffer at pH 2.5 (light colours)

When spectra at pH 2.5 were registered at increasing temperature values from 20°C to 90°C, the protein was found to adopt a random coil structure over 70°C (Tm = 66.1°C), as suggested by the disappearance of the negative band at 215 nm and the presence of a strong negative band at about 200 nm (see Fig. 23A). Spectra deconvolution, performed using the program CDPro [163] (Fig. 23B), showed the increase of a random coil structure with a simultaneous decrease of the β-sheet structure content. The temperature induced transition from β–sheet structure to

random coil at pH 2.5 was found to be reversed when temperature was lowered to 20°C in few minutes (Fig. 24).

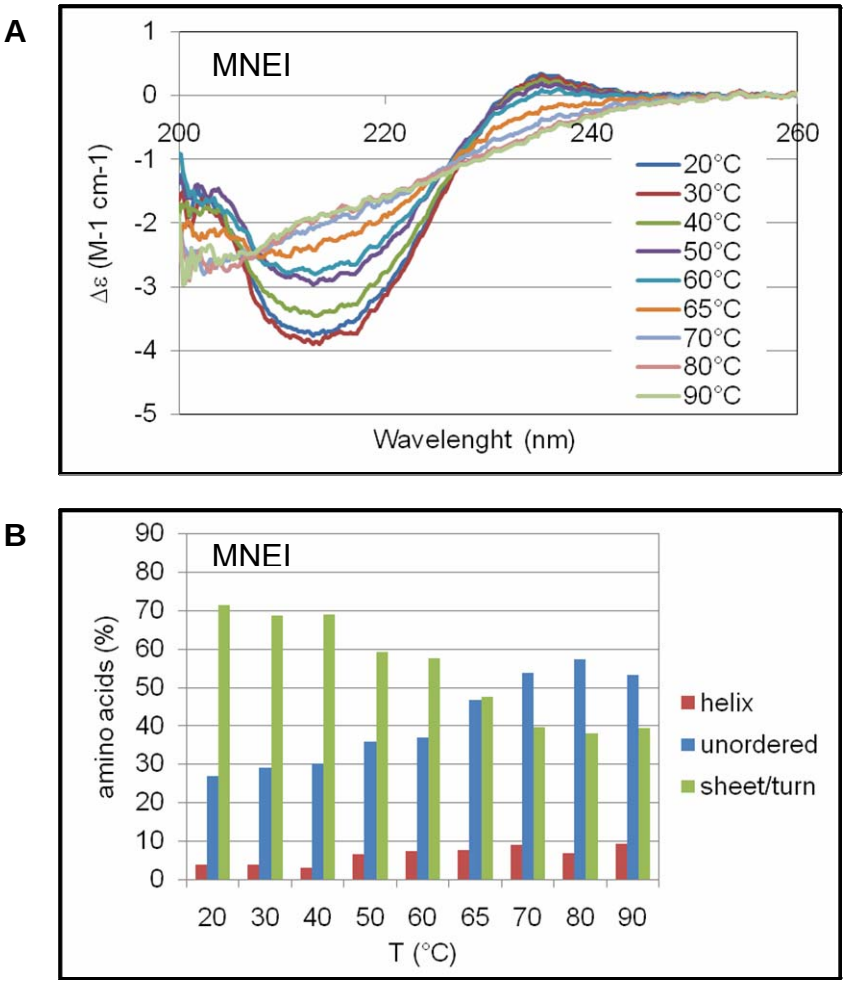


Fig. 23: (A) temperature dependence of CD spectra changes of mnei. CD spectra were recorded at the indicated temperatures in glycine buffer at pH 2.5. (B) CDPro analyses of the spectra. Values are expressed as the percentage of amino acid assuming a particular structure (helix = α-helix; unordered = disordered structure; sheet/turn = β-sheet plus β-turn)

Fig. 25A shows the time dependence of CD spectra changes upon exposure of mnei to high temperature at pH 2.5. We incubated mnei at 86°C, i.e. 20°C above its Tm, and acquired CD spectra after 90 and 150 min of incubation. A decay of the CD signal, associated with protein aggregation and precipitation, was registered.
Furthermore, after 150 min of incubation at 86°C and pH 2.5, the spectra showed the disappearance of the minimum at 200 nm and the presence of a second minimum at about 210 nm, that is suggestive of a transition to a β-sheet structure. when the protein was left at 86°C, a progressive loss of CD signal, due to further aggregation, was monitored by measuring the molar ellipticity at 215 nm in the time interval 90-150 min (Fig. 25B). The decrease of CD signal likely depends on the secondary structure transition, as well as on protein aggregation.

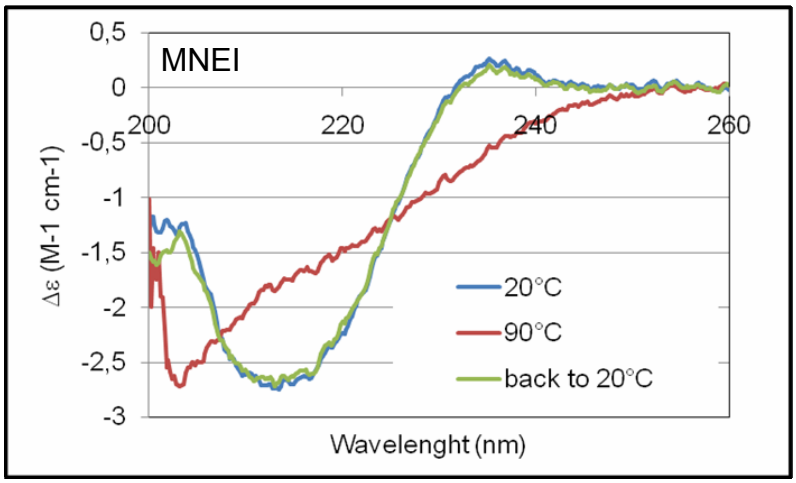


Fig. 24: Reversibility of the temperature-induced transition of mnei in glycine buffer at pH 2.5.

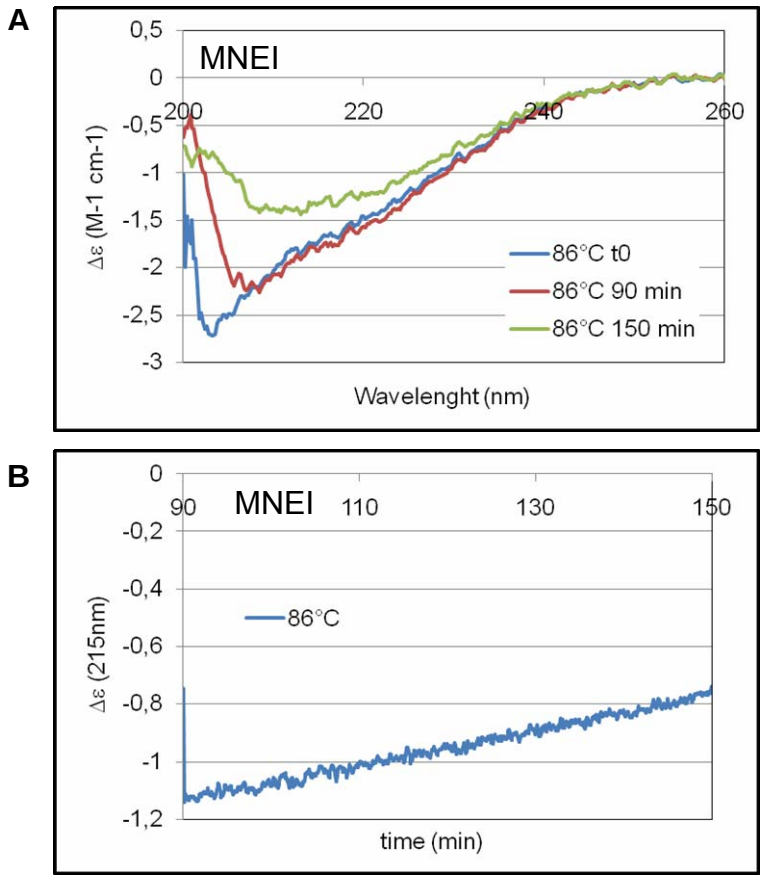


Fig. 25: (A) Time-dependence of the changes in secondary structure content of mnei at pH 2.5. (B) The progressive loss of CD signal was monitored by measuring the molar ellipticity changes at 215 nm during the incubation of protein at 86°C.

The oryc spectra were registered at 20°C, at pH 2.5 and 7.0, the former value being lower than pI, and the latter being higher. The results (see Fig. 22) showed a lower content of β-sheet structure with respect to mnei. This is possibly due to the unfolded structure of the N-terminal segment of oryc, positioned closely to the first β-sheet, and to the differences in the length of the turns. By comparing the spectra at pH 7.0 and 2.5 we concluded that no pH-induced conformational transition occurred. Nevertheless, at pH 2.5 the minimum decreased in intensity with respect to pH 7.0, suggesting that there is a gain in structure content at pH 2.5.
The increase of temperature from 20°C to 98°C at pH 2.5 led the protein to adopt a random coil structure, as suggested by the presence of the strong negative band at about 200 nm (Fig. 26A). Spectra deconvolution with CDPro showed an increase of the unfolded protein content up to 80°C (Fig. 26B). Surprisingly, above 80°C, a recovery in helical content was observed, while the random coil content was found to decrease. As also found for mnei, the temperature induced spectral changes were reversible at pH 2.5 when the temperature was lowered again to 20°C in few minutes (Fig. 27).

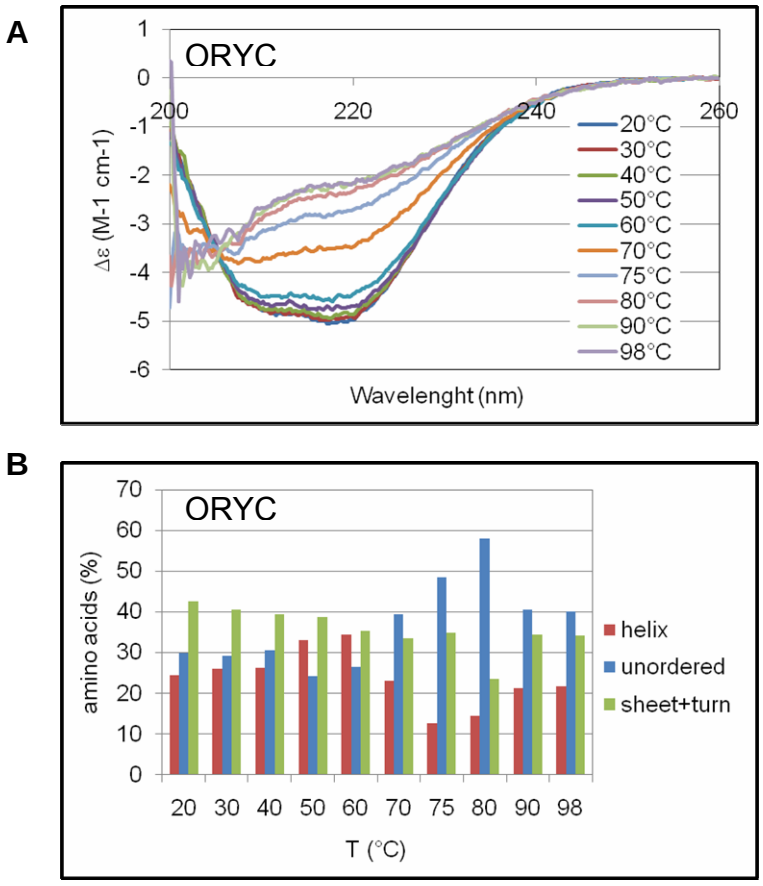


Fig. 26 : (A) Temperature dependence of CD spectra changes of oryc. CD spectra of oryc were recorded at the indicated temperatures in glycine buffer at pH 2.5. A transition from β–sheet to random coil structure occurs at about 80°C. (B) CDPro analyses of the spectra of protein includes also their turn content. Values are defined as in Fig. 23B.

Fig. 28A shows the far-UV CD spectra of oryc, acquired at 90°C (~20°C above Tm) at times 0, 90 and 150 min of incubation in 15 mM glycine buffer pH 2.5, containing 0.15 M sodium chloride.

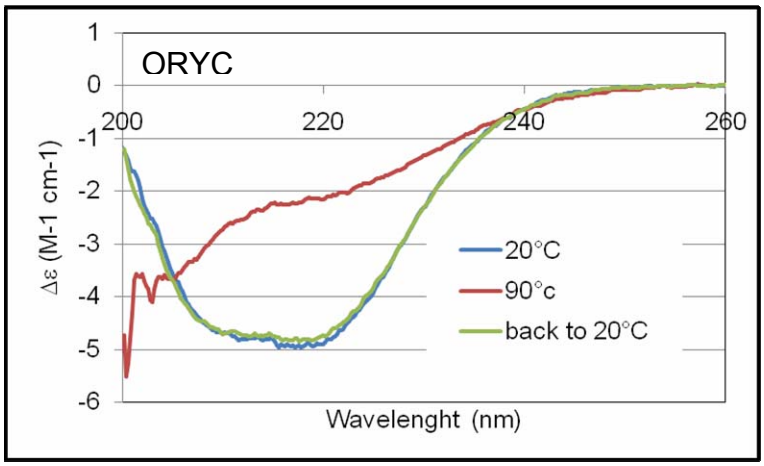


Fig. 27: Reversibility of the temperature induced transition of oryc in glycine buffer at pH 2.5

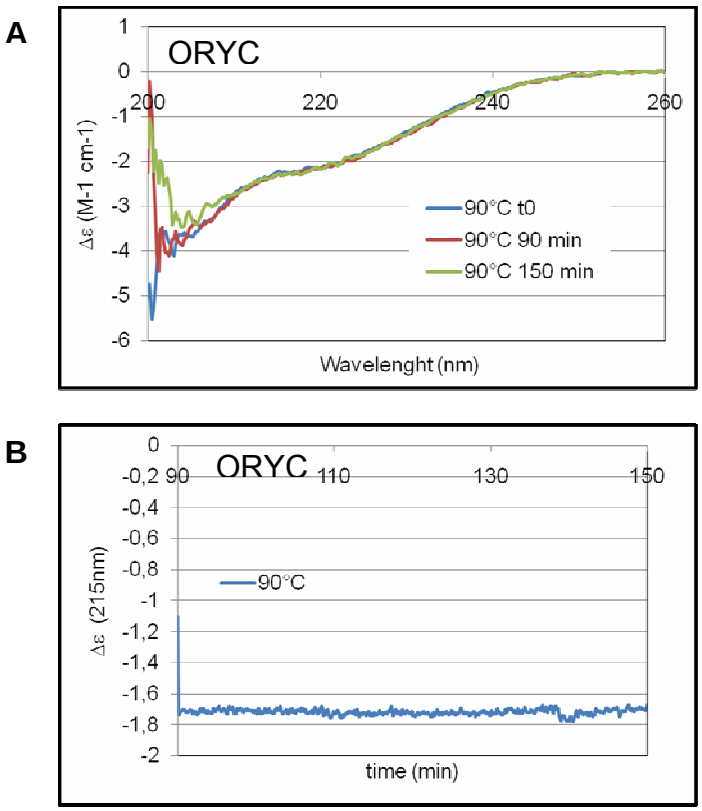


Fig. 28: (A) Time-dependence of the changes in secondary structure content of oryc at pH 2.5. (B) The progressive loss of CD signal was monitored by measuring the molar ellipticity changes at 215 nm during the protein incubation at 90°C.

The spectra were almost superimposable, suggesting the absence of a time-dependent transition, or of phenomena of precipitation/aggregation at the conditions used. Monitoring the molar ellipticity at 215 nm from 90 min to 150 min of incubation in the same conditions, we found the persistence of the CD signal, which suggested that no protein aggregation and precipitation occurred (Fig. 28B).

3.3 Heat-denaturation experiments

To perform heat-denaturation experiments, the protein under test (0.2 mg/ml) was dissolved either in 50 mM phosphate buffer pH 7.0, containing 0.1 M sodium fluoride, or in 15 mM glycine buffer, pH 2.5, containing 0.15 M sodium chloride. The temperature was increased from 20° to 90°C at a rate of 2°C/min. The CD signal at 215 nm was acquired at 1 °C intervals.

Thermal unfolding curves are shown in Fig. 29. The curves show a single transition, indicating that a two-state unfolding process occurred for both mnei and oryc at pH 7.0 and pH 4.0, as expected for proteins of their size (about 10 kDa) [166]. Moreover, thermal unfolding was found to be reversible at pH 2.5 when temperature was lowered again to 20°C (data not shown), consistent with the previous data (Fig. 24 and 27).

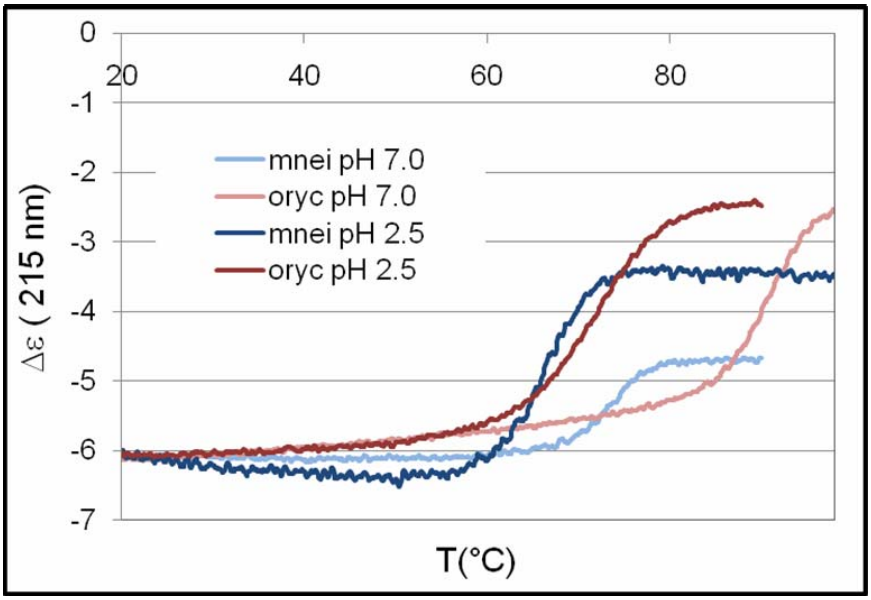


Fig.29: Thermal unfolding curves of mnei (blue lines) and oryc (red lines) in phosphate buffer at pH 7.0 (dark lines) or in glycine buffer at pH 2.5 (bright lines).

To perform a thermal unfolding analysis, the two-state model represented in the Methods section was used.

The parameters obtained from this analysis are shown in Table V. The values indicated that both proteins are less stable at pH 2.5 than at pH 7.0 and that at both pH values mnei is more stable than oryc.

PROTEIN	pH	Tm (°C)	ΔHm (kcal/mol)	ΔG at 20°C (kcal/mol)	Unfolded protein at 20°C (%)
Mnei	7.0	73.7	86.6	7.7	0.0004
OryC	7.0	94.7	76.0	4.8	0.1
Mnei	2.5	66.1	79.6	6.6	0.003
OryC	2.5	73.2	50.1	2.1	5.5

Table V: Thermal unfolding data of mnei and oryc at pH 4.0 and 7.0.

3.4 ThT fluorescence measurements

To verify whether the loss of signal detected in the CD spectra of mnei were to be ascribed to protein aggregation in the presence of fibrilar structures, we performed the ThT fluorescence assay.

As reported above, ThT has a selective response to amyloid fibrillar structures, so this assay is indicative of the presence of amyloid fibrils. Upon binding to amyloid fibrils, ThT undergoes characteristic spectral changes [178] consisting in an increase and in a shift of the major maximum of fluorescence emission from 438 nm to 482 nm [179]. Spectra were recorded in the wavelength interval 455-600 nm upon excitation at 442 nm. Proteins were incubated at 100 μM concentration at 86°C in 15 mM glycine buffer, pH 2.5, containing 0.15 M sodium cloride. ThT fluorescence emission spectra were acquired every 30 min by mixing 7 μL of the protein solution to 800 μL of the ThT solution (20 μM ThT in 25 mM phosphate buffer, pH 7.5). ThT fluorescence emission spectra were acquired immediately after protein addition, in order to avoid perturbation of the fibrillar material due to the transition from acidic pH to pH 7.0.

The spectra of ThT in the presence of mnei showed the typical shift and the increase in ThT fluorescence emission associated to the presence of fibrillar structures, whereas for oryc these characteristic behaviour was not detected (data not shown). In Fig. 30, the kinetics of the variations of fluorescence emission intensity at 482 nm is shown. When mnei was analysed, we found that the increase in ThT fluorescence emission started after about 2 hours of incubation (Fig. 30, green squares).

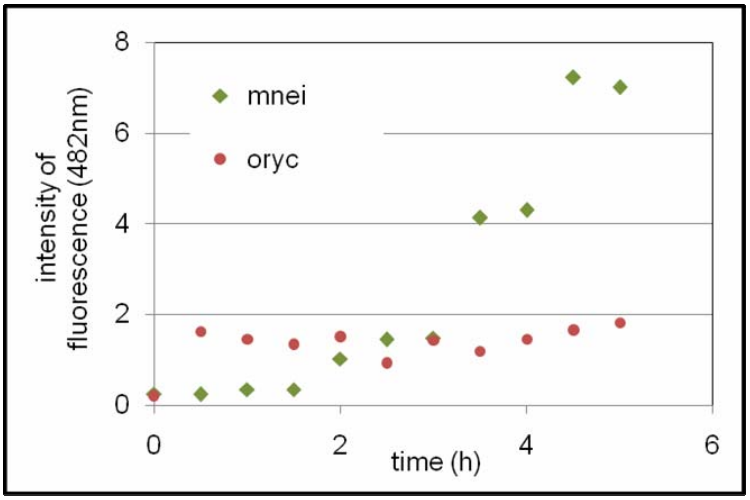


Fig. 30: Kinetics of ThT fluorescence emission at 482 nm during incubation at pH 2.5 of mnei (green squares) or oryc (red dots).

By contrast, ThT fluorescence emission intensity relative to oryc was found to remain constant during the period of analysis (Fig. 30, red circles). This indicated that fibrillogenic structures, generated by mnei and able to bind ThT, appeared during the aggregation process.

3.5 Gel-filtration analyses

Gel-filtration chromatography was performed in order to analyse species size formed during the incubation of monellin and oryc at pH 2.5.

Mnei and oryc proteins were incubated (100 μM) at 86°C in 15 mM glycine buffer, pH 2.5, containing 0.15 M sodium cloride and then analysed by gel-filtration chromatography after 0, 3 and 6 hours of incubation. The results are shown in Fig. 31. Non incubated mnei (t0) was found to be eluted at an elution volume consistent with a monomeric species (Fig. 31A). The same pattern was obtained after 3 and 6 h of incubation. Nevertheless, the peak area corresponding to this monomeric species appeared to be decreased after 3 and 6 h of incubation by about 50% and 80%, respectively. This suggested that during the incubation the soluble monomeric protein was sequestered to form insoluble higher molecular weight species, that were not eluted from the column.

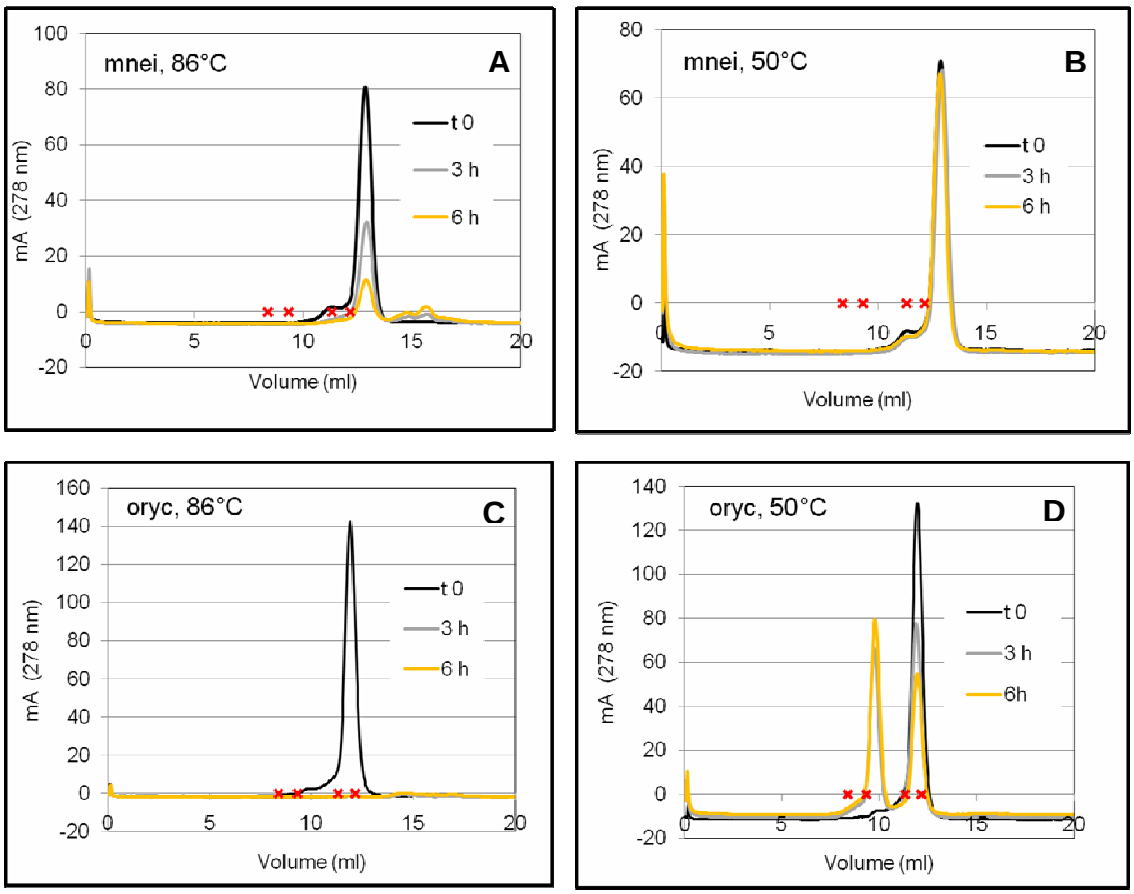


Fig. 31: Gel-filtration analyses of mnei and oryc during incubation at 86°C (A and C, respectively) and 50°C (B and D, respectively). Black lines lines refer to the patterns of non incubated proteins (time 0), grey lines and orange lines refer to the patterns obtained after 3 h and 6 h, respectively. Asterisks correspond to the elution volume of molecular mass standards (from left to right: albumin, 67 kDa; ovalbumin, 49,1 kDa; chymotripsinogen, 20 kDa; RNasi A, 13,7 kDa).

Also oryc protein at time 0 was eluted with an elution volume consistent with a monomeric form (Fig. 31C). Nevertheless, during incubation protein precipitation occurred and no protein was eluted by the column after 6 h incubation. In order to avoid the precipitation of oryc, the incubation temperature was lowered to 50°C. In these conditions, a gel-filtration analysis after 3 and 6 h incubation (Fig. 31D) showed a new species eluted earlier than the monomeric species. For this species, which represents about 40% of the total protein after 3 h, and about 60% after 6 h, a molecular mass of about 36 kDa was calculated.
We hypothesised that the 36 kDa species is a trimeric species and that the formation of this stable oligomer impairs the protein aggregation in amyloid like structures. This hypothesis is in line with the results obtained with the ThT assay. Moreover, the gain of structure of oryc at high temperature and low pH (see CD analyses) confirms the hypothesis on the formation of new occurring species.
Finally, when mnei was incubated at 50°C, no aggregation occurred, as the amount of the monomeric species determined by gel-filtration (Fig. 31C) was found to be unchanged after 3 and 6 h incubation with respect to time 0.

In line with the results reported above, preliminary results obtained by AFM analyses indicated that mnei is able to generate typical fibrils (data not shown).

From our analyses, the following aggregation pathway can be described for mnei and oryc. Both proteins are stable and undergo a two-state unfolding process upon thermal denaturation in acidic conditions. Furthermore, both proteins undergo a reversible and temperature-induced transition from β–sheet to a random coil structure.
Prolonged incubation of mnei under acidic pH triggers the transition to a β–sheet structure, followed by aggregation. Mnei aggregates show the typical amyloid structure, as described by preliminary AFM analyses, and are able to bind ThT.
By contrast, aggregates of oryc generated under the same conditions, do not present any fibrillar structure. Oligomeric species (trimers) were isolated by gel-filtration chromatography. We proposed that these species impair or slow down the amyloid fibril formation of oryc.

4. Catalytic amyloid fibrils

4.1 Designing catalytic amyloid fibrils

Once established the ability of the fibrillogenic polypeptide of ApoA-I to generate ordinate polymeric structures, we planned to exploit the possibility of preparing amyloid fibrils catalytically active, by conjugating the fibrillogenic polypeptide to a suitable enzyme.

Construction of catalytic amyloid fibrils is based on two assumptions. First, the protein chosen as the building block of the amyloid fibrils must be able to aggregate, retaining the typical β–cross amyloid structure, even in the presence of a coupled catalytic moiety. Second, the catalytic activity must be retained even upon protein aggregation in the polymeric amyloid structure.

To prepare catalytic amyloid fibrils by protein engineering, we chose the [1-93]ApoA-I polypeptide as the fibrillogenic moiety and the glutathione S-transferase (GST) enzyme as the catalytic moiety, by fusing them in a chimeric protein, indicated as GST-[1–93]ApoA-I.

GST was chosen as the catalytic partner, as this model enzyme is well characterized both structurally and functionally [182]. Moreover, this enzyme is widely used for biotechnological applications, such as bioremediation and removal of toxic compounds [183]. Nevertheless, technical problems to accomplish this project were to be faced, such as the propensity of GST to generate amorphous aggregates, to irreversibly denaturate upon stressing conditions, and to form covalent multimers in oxidizing conditions.

4.2 Expression and isolation of the fusion protein GST-[1–93]ApoA-I

The chimeric DNA fusion construct, carrying the GST coding sequence positioned upstream to the [1–93]ApoA-I coding sequence, was obtained as previously described. The fusion protein was expressed in *E. coli* BL21(DE3) cells and purified from bacterial extracts using a GSTrap affinity column. SDS-PAGE and Western blotting analyses of the samples obtained by the purification procedure are shown in the paragraph where the isolation of [1–93]ApoA-I is described (Fig. 9 A and B, lanes 1 to 4).

To further purify the fusion protein, anionic exchange chromatography was performed. This procedures was preferred to a reverse-phase chromatography, as the latter was found to irreversibly denature the GST enzyme. The fusion protein was dialyzed against buffer A at pH 8 and loaded on a Resource Q column equilibrated in buffer A, following the procedure described in the Methods section.

The chromatographic profile is shown in Fig. 32. SDS-PAGE and Western blotting analyses of all the fractions eluted by the column (data not shown) indicated that fractions corresponding to peak C contain the expected fusion protein. These fractions were pooled and analysed. In Fig. 33 A, the results of the SDS-PAGE analysis indicated the presence of a single species with the expected molecular mass (36 kDa) of the fusion protein. Western blot analyses (Fig. 33 B and C), performed using anti-ApoA-I antibodies (B) or anti-GST-antibodies (C), indicated that the protein was immunopositive to both antibodies. This confirmed that the expected fusion protein had been isolated as a pure product.

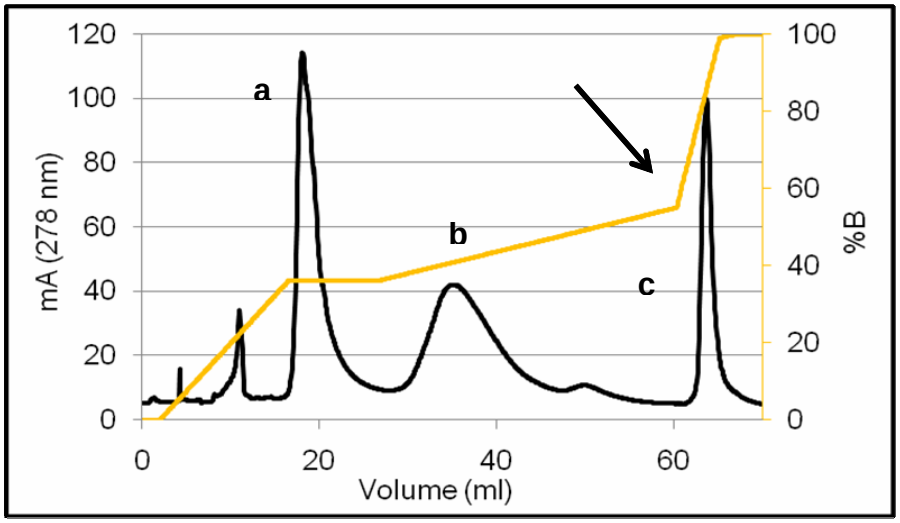


Fig. 32. Ion exchange chromatography profile of the recombinant fusion protein GST-[1–93]ApoA-I. The elution gradient (orange line) and absorbance values at 278 nm (black line) are shown. The arrow indicates the peak corresponding to the fusion protein.

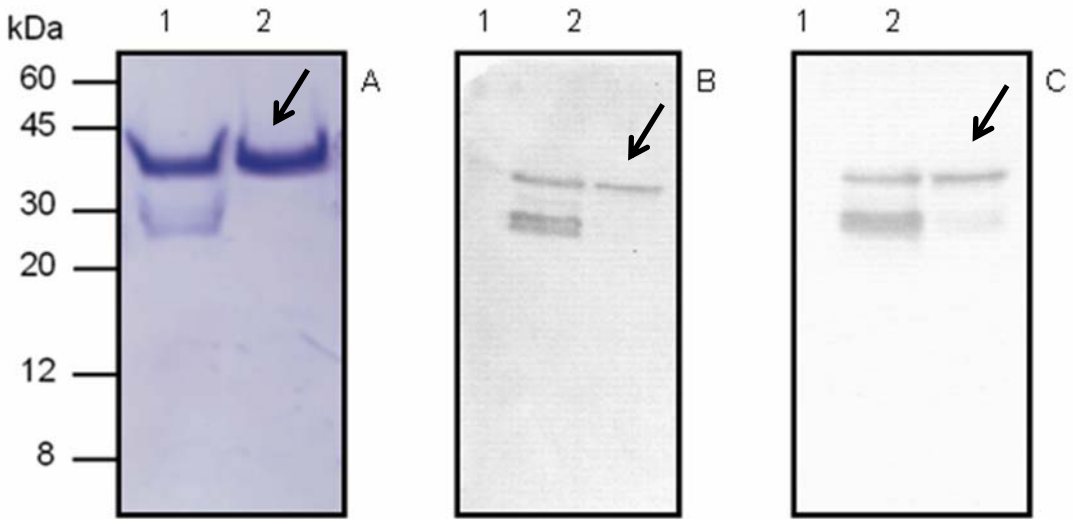


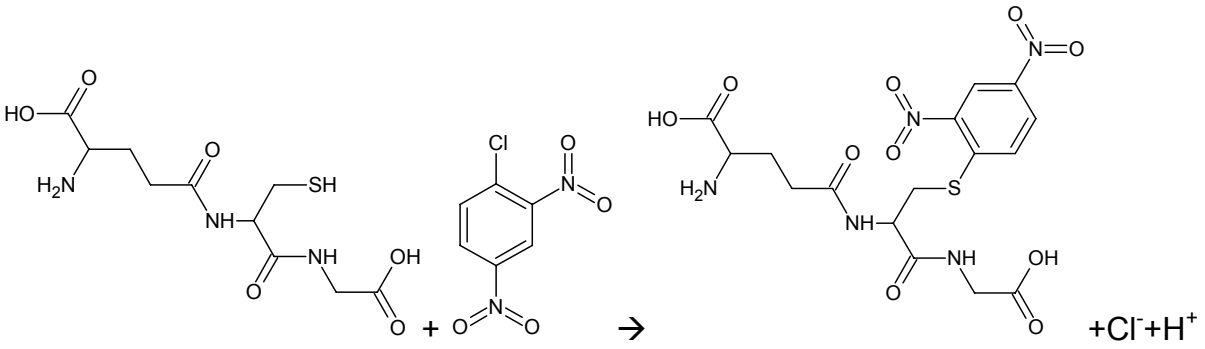
Fig. 33 (A) SDS–PAGE analysis of the recombinant fusion protein GST-[1–93]ApoA-I isolated by ion exchange chromatography. Western blot analyses were performed using anti-ApoA-I-antibodies (B) or anti-GST-antibodies (C). Samples were analysed before (lane 1) and after (lane 2) ion exchange chromatography. The arrows indicate the GST-[1-93]ApoA-I fusion protein.

Using this procedure we were able to purify to homogeneity the fusion GST-[1–93]ApoA-I protein with a recovery of more than 5 mg/L of bacterial culture.
In parallel experiments, we expressed and isolated the GST protein as a recombinant active enzyme to be used as a control in our experiments.

4.3 Characterisation of GST-[1–93]ApoA-I

4.3.1 GST-[1–93]ApoA-I catalitic activity

To test the catalytic activity of the GST moiety of the fusion protein we performed enzymatic assays in the presence of reduced glutathione (GSH) and of the chromogenic compound 1-chloro-2,4-dinitrobenzene (CDNB), commonly used to test GST activity [160]. During the reaction, a dinitrophenyl thioether is produced, that can be measured by the increase of absorbance at 340 nm using a UV/vis spectrophotometer:



As in this reaction a small fraction of glutathione and CDNB react spontaneously, i.e. by a non catalyzed reaction, the corresponding fraction of non-catalytically generated product had to be subtracted (baseline reaction) from the total amount of product obtained.

A typical GST activity assay is shown in Fig. 34, where the activity of the GST enzyme (black line) was measured recording the absorbance variation at 340 nm during the reaction. The assays were performed in 0.1 M potassium phosphate buffer at pH 6.5, following the protocol described in the Methods section. The ΔA value obtained in 1 min with 1μg of enzyme was calculated to be $\Delta A min^{-1}\mu g^{-1}$ = 0.17. The same test, but in the absence of the enzyme was performed to determine the baseline reaction ($\Delta A\ min^{-1}$ = 0.0033). When the fusion protein was tested under the same conditions (Fig. 34 orange line), the GST moiety was found to almost fully retain its enzymatic activity ($\Delta A\ min^{-1}\ \mu g^{-1}$ = 0.16).

These results indicated that the GST enzyme fused to the fibrillogenic peptide is fully active.

4.3.2 Far-UV CD analyses of the fusion protein

A correct, native fold of the GST moiety in the chimeric protein is essential to preserve its catalytic activity. Nevertheless, the presence of the fibrillogenic moiety might alter the correct folding of the GST, therefore impairing its activity. This could be particularly true in the case of [1–93]ApoA-I, known to be a natively unfolded polypeptide.

In order to assess the secondary structure content of the GST-[1–93]ApoA-I fusion protein, we performed far-UV CD analyses of GST-[1–93]ApoA-I. We used as a control a recombinant form, pure and fully active, of GST enzyme. In Fig. 35 CD spectrum of GST (5 μM) at 25°C in buffer A at pH 8 is shown. GST was found to have typical characteristics of an α/β protein, with a predominance of the α-helical

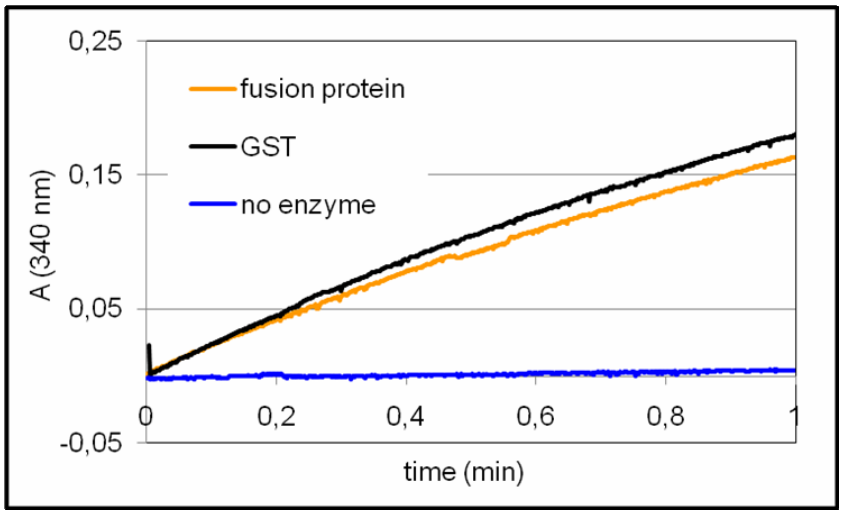


Fig. 34. Analysis of the enzymatic activity of the fusion protein. The chimeric protein (orange line) was tested in the presence of glutathione and CDNB. GST was tested under the same conditions as a control (black line). A non enzymatic reaction was also performed (blue line).

signal (Fig. 35, black line), in agreement with the data riported in the literature [182]. The spectrum of GST-[1–93]ApoA-I performed under the same conditions, was found to be almost superimposable to that of GST, having only a slightly higher content of random coil structure, to be probably ascribed to the presence of the natively unfolded [1–93]ApoA-I moiety (Fig. 35, orange line).

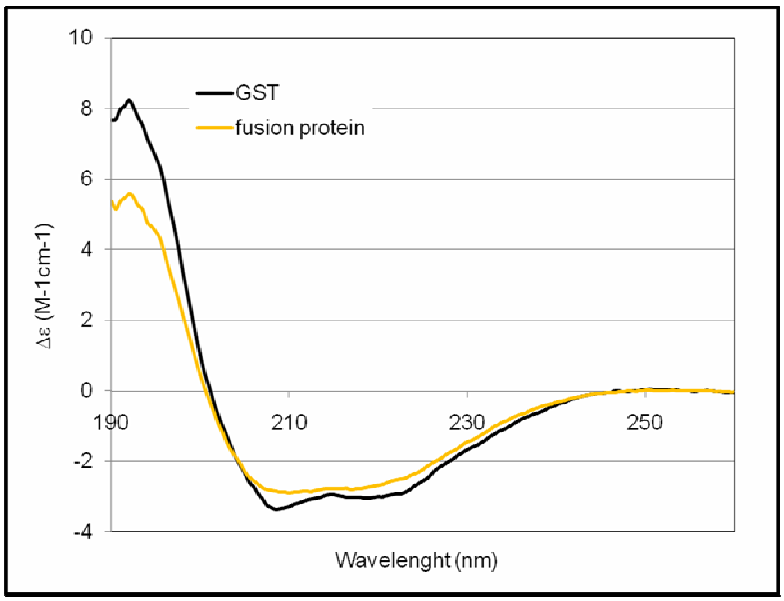


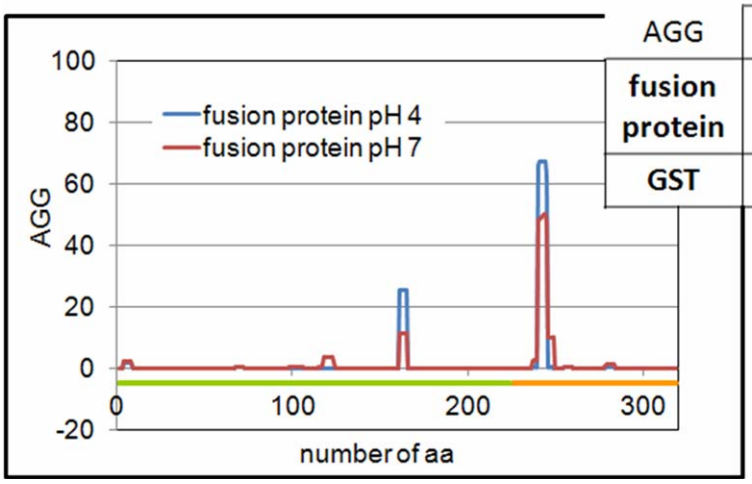
Fig. 35. Far-UV CD spectra of the fusion protein (red line) and of GST (blue line) at pH 8.

Notably, the spectra deconvolution obtained with the CDPro program (Selcon3, CDSSTR, and ContinII methods, data not shown) indicated that the presence of the [1–93]ApoA-I polypeptide at the C-terminus of the chimeric protein did not significantly affect the secondary structure of GST.
These results are consistent with the results obtained on the GST-[1–93]ApoA-I catalitic activity and indicate that the fusion protein retains both the native fold and the catalytic activity of the GST protein.

4.4 Analyses of the fibrillogenic potential of the fusion protein GST-[1–93]ApoA-I

4.4.1 In silico characterization of the GST-[1–93]ApoA-I fusion protein

An *in silico* analysis of the fusion protein was performed in order to predict its aggregation potential. First, the instability index [167] of the fusion protein, as well as that of GST as a control, were analysed and found to be 39.37 and 36.62, respectively. These values indicated that both proteins are stable. On the contrary, the instability index calculated for [1–93]ApoA-I (46.67) indicated that this protein is unstable.
Using the Tango algorithm [62] (Fig. 36) 2 main regions of the fusion protein were predicted to have aggregation propensity at pH 7.0 (red line): a region associated to high aggregation propensity, localised at residues 236-249 of the [1–93]ApoA-I moiety (orange bar), and a region with a minor propensity to aggregation, localised in the GST moiety (residues 161-165). For both regions, the propensity to aggregation was found to be significantly increased when tested at pH 4.0 (blue line).
Thus, the region with the highest propensity to aggregation is localised in the [1–93]ApoA-I moiety, as expected. This analysis also suggested that at both pH values the overall aggregation propensity of the fusion protein is significantly higher than that of GST, clearly due to the presence of the [1–93]ApoA-I moiety (Fig. 36). The overall aggregation values (AGG) calculated for both proteins at pH 7.0 and 4.0 are reported in Fig. 36.



AGG	pH 7.0	pH 4.0
fusion protein	440	535
GST	121	150

Fig 36: Tango analysis of the aggregation propensity of GST-[1–93]ApoA-I at pH 4 (bue line) and pH 7 (red line). The overall agregation propensity (AGG) value for GST and GST-[1–93]ApoA-I at pH 4 and 7 are also reported. The orange bar corresponds to the sequence of [1–93]ApoA-I

4.4.2 Spectroscopic analyses of GST-[1–93]ApoA-I fusion protein

In order to analyse the ability of the fusion protein to undergo conformational transitions and to form amyloid fibrils, we performed spectroscopic analyses using specific dyes such as ANS, ThT and CR. Recombinant GST was used as a control. The aim of these analyses was to verify whether the chimeric protein retained the spectroscopic characteristics of a fibrillogenic protein.

4.4.3 ANS binding

ANS, a very common fluorescent probe for protein studies, has been used to monitor protein conformational changes, as it binds to exposed hydrophobic regions of proteins. As a consequence, when a conformational transition occurs, the fluorescence intensity of ANS increases and the emission spectrum maximum shifts from 520 nm to 470 nm [17].

Binding of ANS to the fusion protein or to GST, induced by pH acidification, is shown in Fig. 37 A and B, respectively. ANS emission fuorescence spectra were recorded in the range of 400–600 nm at the excitation wavelength of 380 nm.

The results indicated that: (i) for both proteins spectra at pH 8.0 were found to be almost identical to the spectrum of ANS dye; (ii) when the pH value was lowered to 6.4, for both proteins a blue shift and a significant increase in the maximum of fluorescence emission were detected; (iii) when the pH value was returned to 8.0, for both proteins the spectral changes were reverted.

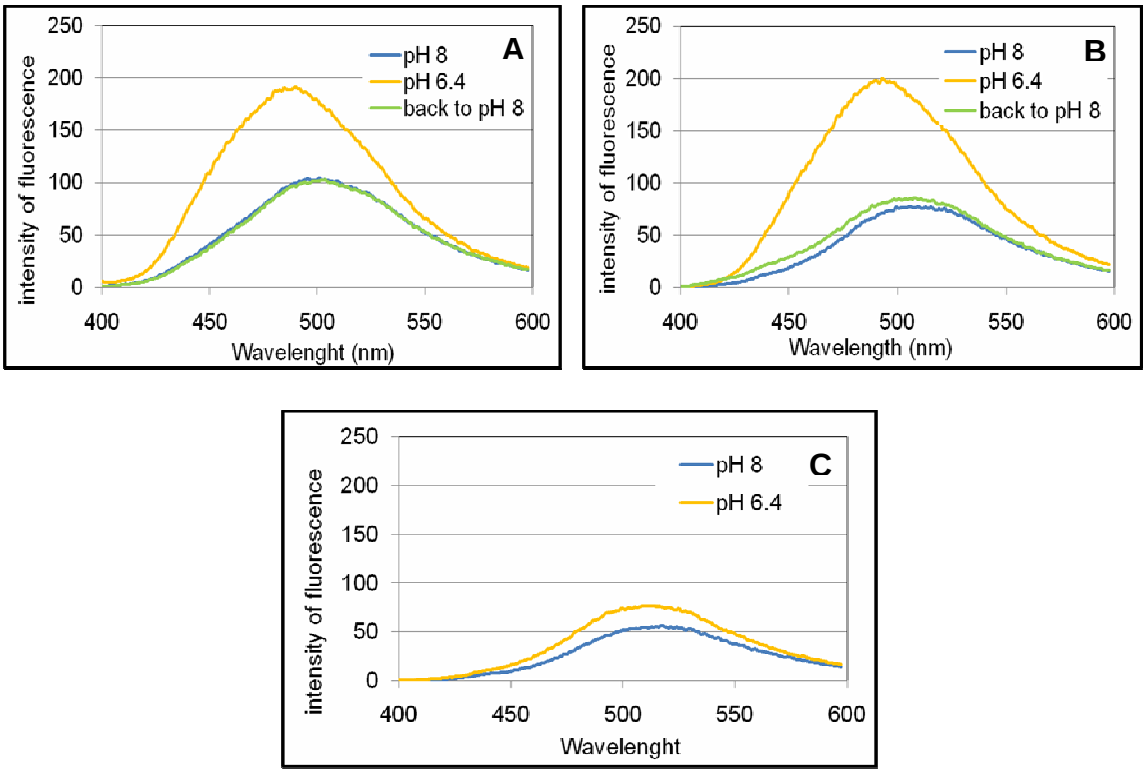


Fig. 37: pH-induced binding of ANS to fusion protein (A) or GST (B). ANS spectra at pH 8.0 are shown in blue whereas spectra at pH 6.4 are shown in orange. In green are shown the spectra registered upon returning to pH 8.0. Spectra of the dye are shown (C) at pH 8.0 (green) and 6.4 (orange).

In conclusion, according with CD analyses, both proteins at pH 8.0 are correctly folded; no hydrophobic regions are exposed to the solvent. Upon acidification, a pH induced conformational transition occurs for both proteins, which expose to the solvent hydrophobic regions previously buried in the protein core. This transition is frequently found in the early step of the fibrillogenic pathway of proteins [175] and it has been related to a molten globule state. However, no spectral differences were detected between the GST and fusion protein. In the literature, fluorescence analyses of the early unfolding process of GST have been reported that confirm our results [184].

4.4.4 Thioflavin T assay

As the pH induced conformational transition is known to be an early event occurring in the fibrillogenic pathway of some amyloidogenic protein [38], we searched for the presence of amyloid structure generated during long time protein incubation in buffer A at pH 6, using the ThT dye binding assay. As previously reported, Thioflavin T has been widely used in the *in vitro* fluorometric analyses to detect amyloid fibrils, as it associates rapidly with the aggregated fibrils. Upon binding, the excitation and emission maxima were shifted from 385 nm and 445 nm, respectively, to 450 nm and 482 nm, respectively.

Proteins (10 μM) were incubated at 25°C in buffer A in the presence of ThT 15 μM. The increase of the ThT emission fluorescence at 482 nm (excitation 450 nm) was recorded using a spectrofluorimeter.

In Fig. 38A the results of our experiments are reported. For the fusion protein (blue circles), a time dependent increase of emission fluorescence was detected. The obtained curve is typical for amyloid polymerisation kinetics, with a short lag phase (about 3 h). These results indicated that during GST-[1–93]ApoA-I incubation, amyloid-like structures were generated, able to bind the ThT dye.

By contrast, ThT fluorescence emission did not change significantly during the incubation of GST under the same conditions (red squares Fig 38A), suggesting that no amyloid-like structures occurred during incubation.

When the aggregation curve of GST-[1–93]ApoA-I was compared to the aggregation curve of the fibrillogenic polypeptide [1–93]ApoA-I (see Fig. 15), the conclusion could be drawn that the presence of GST increased the fibrillogenic potential of [1–93]ApoA-I polypeptide. In fact, it is well known that the protein context plays an important role in determining the stability and solubility of fibrillogenic polypeptides and may modulate the aggregation properties of amyloidogenic fragments [185].

Following these experiments, SDS-PAGE and *Western blot* analyses under reducing conditions were performed on GST-[1–93]ApoA-I samples. Besides the presence of monomeric GST[1–93]ApoA-I species, the presence of species with high molecular weight (Fig 38B) was detected and found to be immunopositive to anti-ApoA-I antibodies (Fig.38C). The fact that these species were detected under denaturing and reducing conditions, suggested that they are very stable, although non covalently linked. As a consequence, we might exclude that aggregation by disulphide bridges occurs as sometimes reported in the literature for GST [186].

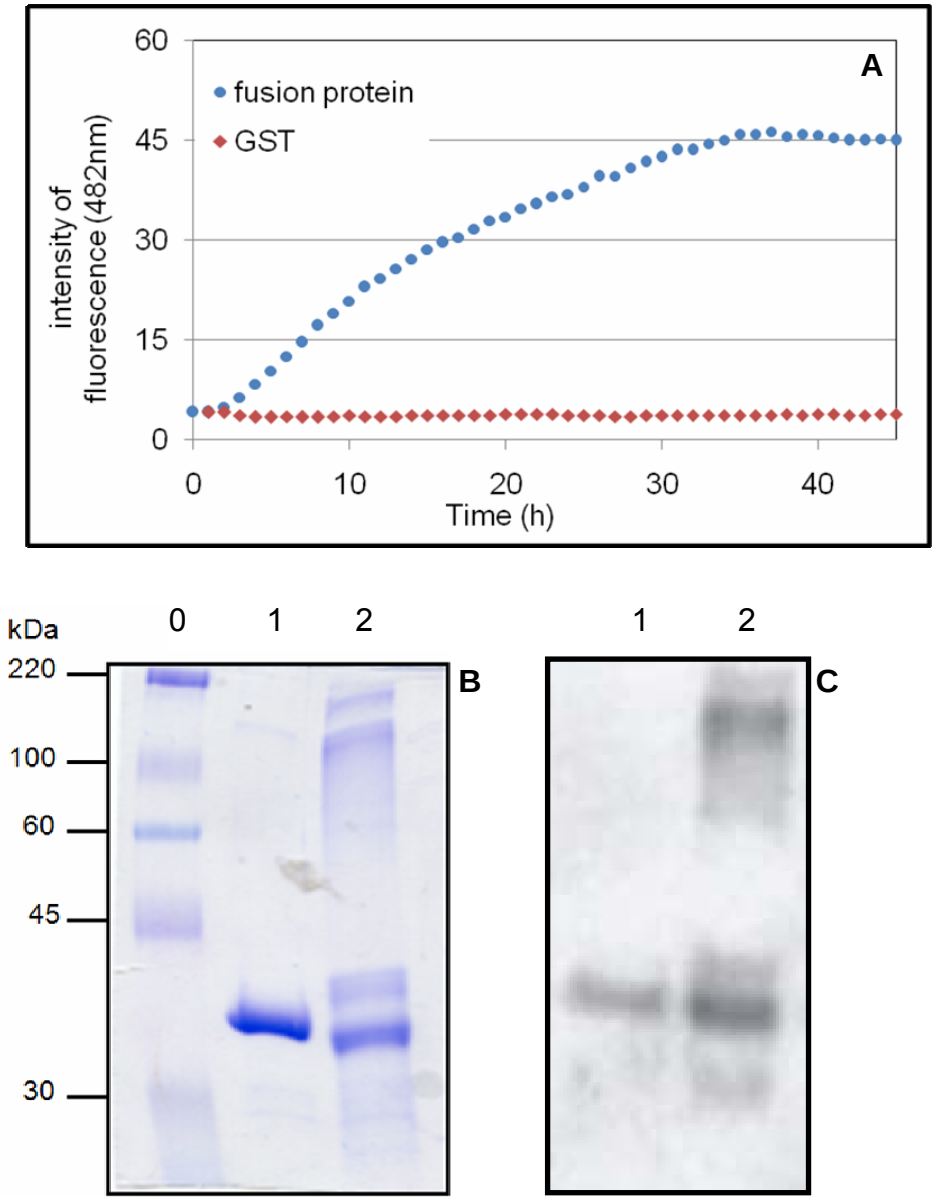


Fig. 38: (A) The increase of ThT emission fluorescence at 482 nm during the incubation at 25°C of the fusion protein (blue circles) or GST protein (red squares). (B) SDS-PAGEand (C) *Western blot* analyses of the fusion protein tested at the end of the experiment reported in (A). Lane 0, molecular weight standards, Lane 1, the non incubated protein; lane 2, the incubated protein.

4.4.5 *Congo Red analyses*

To further characterise the putative amyloid-like structure of GST-[1–93]ApoA-I, we also performed Congo red assays. CR is believed to specifically interact in some unknown way with the crossed-β-sheet structure common to amyloid structures. As previously described, the Congo red assay is a differential spectroscopic assay, in which the Congo Red bound fibrils is measured by means of a typical spectral red shift of the adsorption maximum, from 490 nm to 540 nm in the spectrum of the Congo red-protein complex. GST-[1–93]ApoA-I, or GST as a control, were incubated in the same conditions described for ThT assays. After about 60 h of incubation, CR

assays were performed by diluting each protein to 100 μg/ml followed by the addition of Congo Red dye (20 μM) [176].
The results are shown in Fig. 39. By subtracting both the spectrum of GST-[1–93]ApoA-I alone (green line, Fig. 39 A), indicative of the light scattering phenomen, and that of Congo red alone (blue line, Fig. 39 A) from the spectrum of GST-[1–93] ApoA-I in the presence of Congo Red (orange line, Fig. 39 A), a differential spectrum was obtained (Fig. 39 C, red line). This spectrum showed the typical maximum red shift indicative of Congo Red binding to fibrillar structure. This suggested the presence of polymeric amyloid β-cross structures, in line with the data obtained with ThT.
When the same experiments were performed with GST (Fig. 39B), no red shift was detected, indicating the absence of amyloid structures, in line with the results, obtained with ThT.

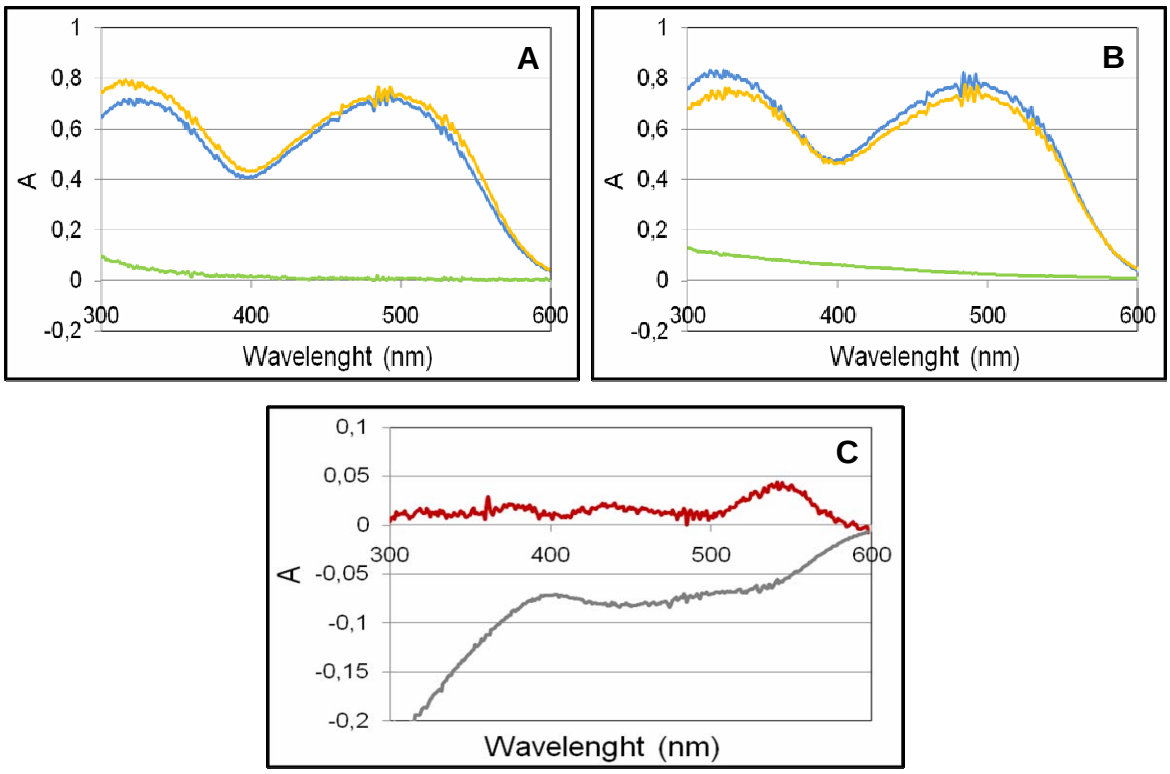


Fig. 39: Congo Red spectra in the presence of the fusion protein (A) or GST (B). The spectra of the proteins in the presence of Congo Red are in orange; the spectra obtained with proteins alone are in green; those of Congo Red are in blue. (C) The differential spectra of the complexes fusion protein-Congo Red (red line) or GST-Congo Red (gray line), obtained by subtracting the green and blue spectra from the corresponding orange spectrum.

Taken together, the spectroscopic data collected using different methodologies (CD, ANS, ThT, CR) strongly suggested that: (i) for both fusion protein and GST, a molten globule state is induced by lowering the pH; (ii) the chimeric protein GST-[1–93]ApoA-I is able to evolve towards a fibrillogenic, amyloid-like structure; (iii) GST being unable to polymerize, generates amorphous aggregates.

4.4.6 Morphological analysis of fusion protein aggregates

The fusion protein GST-[1-93]ApoA-I (0.2 mg/ml) was incubated under conditions that promote fibrillogenesis, i.e. 1 week at 25°C, at pH 6.4. Following incubation, the insoluble protein fraction, recoved by centrifugation, was analysed by AFM by the group of Prof. A. Gliozzi, University of Genova on mica surfaces.

Typical fbrillar structures were detected. In Fig. 40, protofilaments generated by the fusion protein self-assembly, are shown. The presence of heterogeneous aggregated material was also described (not shown).

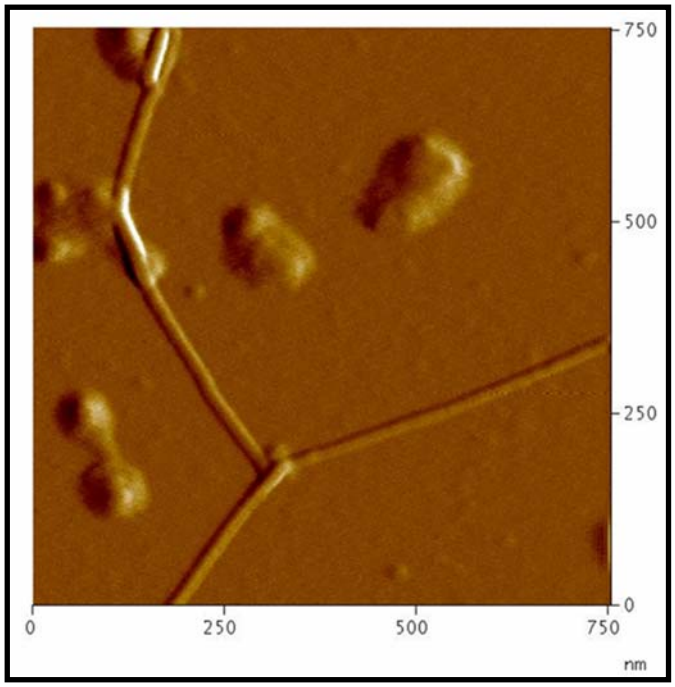


Fig. 40: AFM images fibrils generated by the fusion protein.

4.5 Analysis of the catalytic activity of GST-[1–93]ApoA-I fibrils

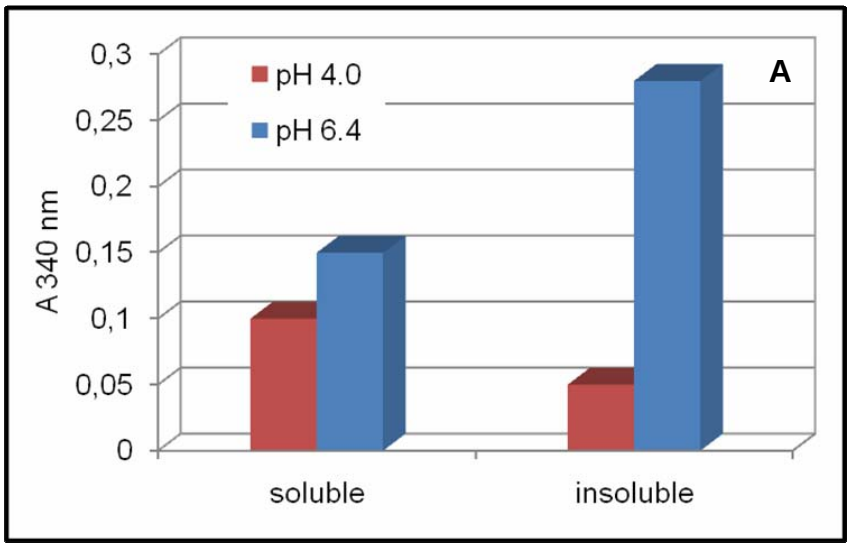
Once verified that the fusion protein in its soluble state retained the GST catalytic activity, and that it was able to form amyloid-like structures, we tested the catalytic activity of these fibrils.

Following the procedure to induce fibril formation (i.e. incubation for at least 1 we3ek at pH 6.4), the protein was dried and resuspended in buffer A at pH 6.5. Soluble and insoluble fractions were separated by centrifugation and tested by the CDNB assay.

The results are reported in Fig. 41A, where the increase of absorbance at 340 nm was determined after 45 min of reaction. The insoluble material, generated during the incubation at pH 6.4, was found to be enzymatically active, whereas low levels of activity were found to be associated to the corresponding soluble fraction. No activity was instead found to be associated to the aggregated material generated at pH 4.0, which suggested that in acidic conditions amorphous, inactive aggregates were prevalently produced.

The same samples were also analysed by SDS-PAGE under reducing conditions. This analysis showed that high molecular weight species were present in the protein

incubated at pH 6.4 (Fig. 41B). These specie were not detectable in the non incubated protein, or in the protein incubated at pH 4.0.
As the SDS-PAGE analysis was performed under reducing conditions, we might exclude that the insoluble material, enzymatically active, was generated by oxidative reactions, such as S-S bridges formation. In fact, it is known that GST under non reducing conditions generates covalently linked aggregates without loss of enzymatic activity.



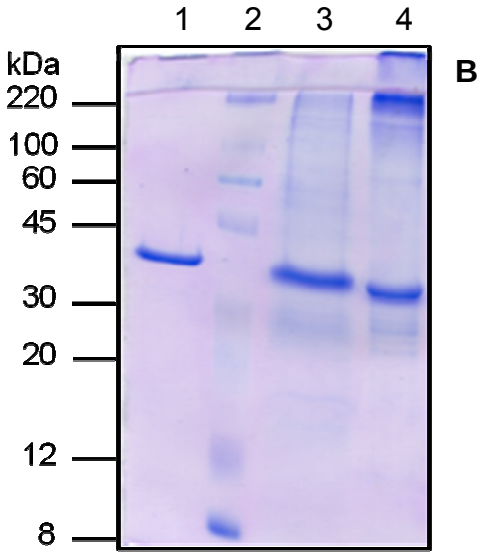


Fig. 41: (A) CDNB assay of the chimeric protein aggregates generated after 7 days incubation at pH 6.4 (blue bars) or pH 4.0 (red bars). The soluble and insoluble protein fractions were tested. (B) SDS-PAGE analysis in reducing conditions of the chimeric protein aggregates. In this case, the total protein was analysed without separating soluble from insoluble fraction. Lane 1, GST-[1–93]ApoA-I; lane 2, molecular weight standards; Lane 3, aggregates of GST-[1–93]ApoA-I generated at pH 4.0; Lane 4, aggregates generated at pH 6.4 (lane 4).

These results are in line with the data collected by ThT.
The hypothesis can be raised that, whereas at pH 4.0 most of the protein precipitates in amorphous aggregates with loss of catalytic activity, at pH 6.4, typical amyloid fibrils are generated. These ordered structures may thus be viewed as a new enzyme linked polymeric support. Thus, a new catalytic matrix was produced starting from fibrillogenic polypeptide.

4.6 Trapping the catalytic amyloid fibrils onto a solid support

To immobilise the catalytic amyloid fibrils on a solid support, such as filters, PVDF membranes were chosen with a cut-off of 0.02 or 0.2 μm. Filters were not activated, as usually recommended by the manufacturers, to avoid protein binding and allowing filters to trap exclusively high molecular weight species. A suspension of fibrillar species (1 ml) was filtered through the two types of membranes; the filters were then washed with 5 ml of buffer in order to remove low molecular weight species.
CDNB assays were then performed to test the enzymatic activity of the fibrillar species immobilised on filters, as well as that of the protein fraction washed out from the filters.
Then, the enzymatic activity associated to the fibrils immobilised onto 0.02 μm or 0.2 μm filters was tested in parallel experiments. To perform these assays, each filter, positioned within its plastic holder, was connected to a peristaltic pump. The reaction mixture, containing the substrates of the GST enzyme, was let to pass continuously through the filter for 10 min. The enzymatic activity was expressed as the increase of the absorbance recorded after 10 min of reaction.
The results of these experiments are shown in Fig. 42.

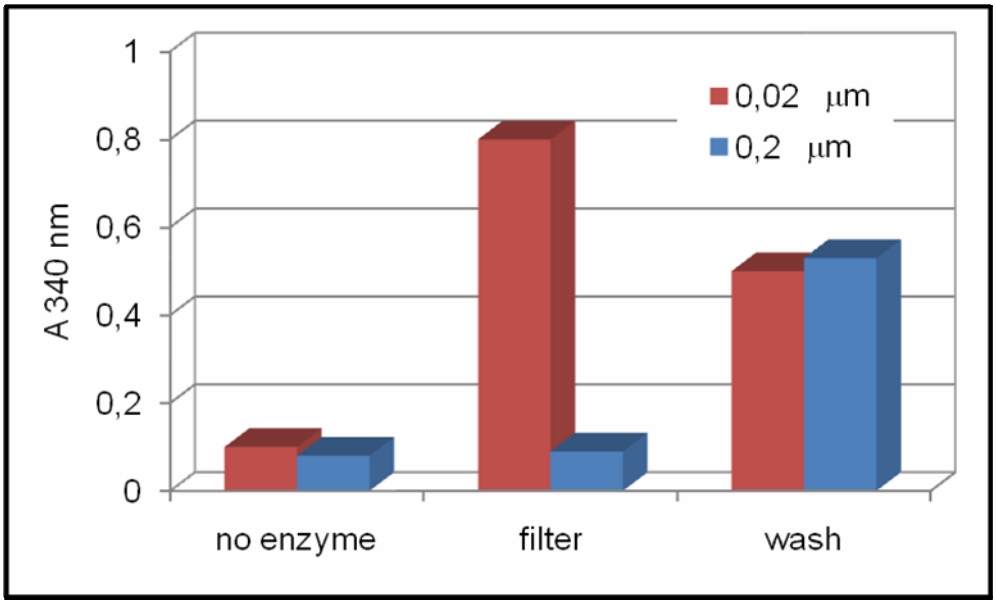


Fig. 42: Catalytic fibrils were loaded on filters with cut-off of 0.02 or 0.2 μm and then washed with buffer. After 10 min of CDNB reaction, the absorbance variations were determined for both samples, as well as for the washed out material.

As shown in the Figure, fibrils immobilised on filters with cut-off 0.02 μm were found to be enzymatically active, i.e. able to generate the expected reaction product. No activity was found instead to be associated to fibrils loaded on filters with 0.2 μm cut-off.
This indicated that the size of the fusion protein fibrils were too small to be trapped on filters with cut-off 0.2 μm, but sufficiently large to be trapped by 0.02 μm cut-off filters.
Some enzymatic activity was also detected in the protein fraction washed out by the filters, indicating that some low molecular weight species, possibly small oligomers, were also present in the fibrillar material.
These results indicated that fibrils generated by the chimeric protein are costituited by self-assembled high molecular weight polymers, endowed with catalytic activity.

DISCUSSION

1. The recombinant fibrillogenic polypeptide [1-93]ApoA-I

Amyloidogenic proteins and peptides have high propensity to aggregate in fibrillar structures [49]. The presence of specific mutations, and/or environmental conditions, that tend to destabilize the native conformation of a fibrillogenic proteins, represents the principal cause that may kinetically favour aggregate nucleation. In amyloid diseases associated to apolipoprotein A-I (ApoA-I), the major component of high density lipoproteins (HDL), the pathogenesis is associated to the presence of mutated versions of the 243-residue native protein [84]. Fibrils extracted from amyloid deposits were found to be mainly constituted by N-terminal fragments of ApoA-I, about 90-100 residue long [101]. In particular, the polypeptide 1-93 was found to be the main constituent of cardiac fibrils extracted from patients affected by amyloidosis, harbouring the mutation L174S in the full length protein.

The 93-residue ApoA-I polypeptide, extracted from ex-vivo fibrils, is a "natively unfolded protein" at neutral pH, but at pH 4 it undergoes a complex fibrillogenic pathway [102]. The difficulty to obtain the natural polypeptide from ex-vivo sources in sufficient amount made so far impossible further structural and functional studies. Despite the instability of the fibrillogenic polypeptide, we succeeded in the production and isolation to homogeneity of a stable recombinant form of the 93-residue fibrillogenic domain of ApoA-I, denoted as [1-93]ApoA-I [174].

In this work we cloned, expressed and purified the [1-93]ApoA-I polypeptide using an effective and reliable expression system to produce the recombinant form of the amyloidogenic polypeptide.

The recombinant 1-93 ApoA-I polypeptide was expressed in bacterial cells following an experimental strategy aimed at reducing the intracellular degradation during its production. Briefly, the polypeptide was expressed as a chimeric product, obtained by fusing the polypeptide to a stable bacterial protein. The polypeptide was then released by site-directed proteolysis and isolated as a stable, pure protein.

The final product was analyzed to investigate its conformational properties and fibrillogenic potential, using a multidisciplinary approach.

Conformational dynamics of the recombinant [1-93]ApoA-I was inspected by CD and fluorescence analyses. A conformational switch from a random coil to a molten globule/helical conformer, able to bind ANS, was clearly detected by lowering the pH from 7 to 4 [174], as it has been described for the natural polypeptide [102]. The transition was found to be fully reversible when the pH is returned to 7.

We proposed that the helical conformer is a key intermediate in the fibril formation pathway and that its generation represents an early and reversible event in the fibrillogenic pathway.

The helical/molten globule intermediate displayed a strong propensity to oligomerize. During a prolonged incubation at pH 4.0 it was found to switch towards a predominant β-sheet based structure that evolves to an insoluble polymeric structure.

The experimental data obtained from conformational analyses are in line with the *in silico* analysis of the aggregation propensity of [1-93]ApoA-I, performed with the most common predicting algorithm available (Tango algorithm). The polypeptide was predicted to have a high and pH-dependent aggregation propensity.

Further characterisation of the aggregated material was performed using amyloid specific dyes, as thioflavin T (ThT) and Congo Red (CR). These analyses indicated that the aggregated material is organized in regular and ordered polymeric amyloid

structures. In these assays, the acidification emerged as a condition able to increase the rate of amyloid formation, as found for the most common amyloidogenic proteins [38].
Gel-filtration analyses of [1-93]ApoA-I showed that the polypeptide is a dimer under physiological conditions. Incubation of the polypeptide under acidic conditions, followed by the gel-filtration analyses, showed the disappearance of the dimeric species, presumably sequestered in insoluble oligomers. Interestingly, the existence of dimeric species had been also suggested by previous experiments, performed by Prof. P. Pucci of the University of Naples, using limited proteolysis coupled with mass spectrometry.
The data collected in the laboratory of Prof. A. Gliozzi of the University of Genova, which collaborates with our group as an expert of microscopy techniques, showed that the recombinant polypeptide, analysed by Atomic Force Microscopy, is able to generate fibrils in a time scale comparable to that of its natural counterpart. Upon incubation of the polypeptide in acidic conditions, narrow protofilaments and well-defined fibrils were detected together with globular aggregates.
From our data, a complex and multistep intriguing aggregation pathway emerges for [1–93]ApoA-I. Major conformational rearrangements are involved and different intermediate species with own structural characteristics are generated during the aggregation process, in complete agreement with the data obtained for the natural peptide [102]. Our analyses strongly suggest that a delicate balance between natively stabilizing interactions and pathological propensity to self-aggregation exists, stressing the importance of environmental conditions in the maintenance of protein conformation.

2. Towards the understanding of the molecular bases of [1-93]ApoA-I aggregation

The role of mutations in increasing the propensity of proteins to aggregate is a crucial aspect in the ApoA-I amyloidoses, where the disease is invariably associated to the presence of specific mutations in ApoA-I gene.
All the patients analysed so far were found to be heterozygous for the mutated protein. Nevertheless, the observation that, when an internal mutation occurs, i.e. a mutation located in the N-terminal domain of ApoA-I, almost only the polypeptide carrying the mutation was detected in fibrils. This suggests a role of amyloidogenic mutations in modulate the aggregation propensity of [1-93]ApoA-I polypeptides.
As discussed above, 8 mutations in the N-terminal domain of ApoA-I have been demonstrated so far to be associated to amyloid pathologies.
The availability of a suitable protocol to express [1-93]ApoA-I encouraged the production of its amyloidogenic variants. This will open the way to a detailed comparative study on the conformational dynamics and fibrillogenic potential of the mutated forms versus the wild type polypeptide, aimed at the comprehension of the molecular mechanism by which mutations in ApoA-I promote amyloid deposition.
Following the protocol set up for [1-93]ApoA-I, we were able to express and isolate to homogeneity the 8 mutant versions of [1-93]ApoA-I polypeptide, thus generating the complete panel of variants of the 93-residue ApoA-I fibrillogenic polypeptides (Guglielmi et al., manuscript in preparation).
Six of them harbour a single residue substitution (mutants G26R, T50R, L60R, L64P, L75P and L90P), whereas two mutants harbour a deletion at sequence 70-72 and 60-71, respectively (in the latter case the deleted sequence is replaced by VT) [84].

Experiments aimed at the characterization of the conformational state and structural dynamics of the 8 amyloidogenic variants of [1–93]ApoA-I are in progress. Preliminary data, obtained in collaboration with Prof. V. Bellotti of University of Pavia, indicated that almost all mutants display an increased aggregation propensity with respect to wild type [1-93]ApoA-I.
The availability of recombinant forms of wild type [1-93]ApoA-I and its mutated versions represents a powerful *in vitro* system to perform a systematic study on the effects of each mutation on the fibrillogenic potential of these polypeptides. Further investigation of [1–93]ApoA-I polypeptides, using independent experimental approaches, such as CD, fluorescence and microscopy, will integrate our preliminary results, with the aim to define a suitable algorithm able to correlate structural determinants to functional features.
In the next future, our molecular system will be used to analyse the molecular bases of ApoA-I amyloidosis. In this regard, the understanding of the mechanism of amyloidosis represents an important step towards the designing of therapeutic approaches aimed at the treatment or prevention of amyloid disease

3. Plant cystatins as model proteins to study the fibrillogenic process

Plant cystatins are protease inhibitors that represent an interesting model system to shed light on cystatins aggregation behavior. These proteins were found to be able to form amyloid fibrils under specific conditions [124]. Plant cystatins share structural features (the cystatin motif) with human cystatins, which are involved in human severe amylopathies. Among them, human cystatin C is responsible the hereditary cystatin C amyloid angiopathy, named HCCAA [115], whereas human cystatin B is associated to Unverricht- Lundborg disease [117].
In the period June-September 2007, I worked at the Medical Research Council (MRC), London, National Institute for Medical Research, in the laboratory of Dr. A. Pastore. My research project concerned studies on the conformational and aggregation properties of two plant cystatin: mnei, a recombinant, sweet tasting monellin, and oryzacystatin I (oryc), a cystatin from *Oryza sativa*.
When tested by thermal unfolding analyses, both proteins were found to be very stable even at acidic pH. As inferred by conformational analyses performed by CD, we found that under acidic conditions these proteins undergo a β–sheet to random coil transition, induced by increasing the temperature. Longer exposure of mnei in acidic conditions causes the appearance of a β–sheet signal, that is lost upon prolonged incubation due to protein aggregation and precipitation.
Although both proteins were found to aggregate, the aggregated material generated by mnei was able to bind amyloid specific dyes, such as thioflavin T, whereas oryc aggregates did not assume amyloid-like structure. Moreover, gel-filtration analyses of mnei, after incubation at high temperature, showed the disappearance of the monomeric species, reasonably sequestered in protein multimers.
Interestingly, the time scale of the appearance of mnei β-cross structure, inferred from ThT assays, is compatible with the time scale of the conformational transition and protein aggregation inspected by CD analyses.
On the contrary, oryc was found to be able to generate stable trimers, as indicated by gel-filtration analyses. To explain the absence of amyloid structures in oryc aggregates, we hypothesised that these trimers are able to prevent the protein from self-aggregation in amyloid structures.

Actually, the presence of dimeric or tetrameric forms of human cystatin, in which the subunits are linked together by domain swapping, were demontrated by crystallographic and gel- filtration analyses [119, 12].
Hence, the formation of stable oligomers by domain swapping could be a general mechanism to prevent amyloid fibrillogenesis.

4. A biotechnological application of amyloid self-assembly

Molecules may undergo self-association forming hierarchical structures. Self-assembly can be defined as the spontaneous organization of individual components into an ordered, well-defined and organized structure. Molecular self-assembly ("molecular Lego"), by definition, is the spontaneous organization of molecules under thermodynamic equilibrium conditions into structurally well-defined and rather stable arrangements through a noncovalent interactions [11]. Molecular self-assembly is emerging as a new route to produce novel materials and to complement other materials.
Amyloid fibrils formed by self-association of polypeptide chains are therefore possible candidates for the construction of nanostructured materials, using the bottom-up strategy, from single molecules to sopramolecular structures. Amyloid fibrils are of interest in the field of protein engineering by means of their well-ordered and multidimensional structure [187], the possibility to control the assembly conditions, and the presence of weak noncovalent interactions rather than covalent bonds.
We inspected the possibility to generate new catalytic matrices using the amyloid fibril as physical supports on which to align enzyme molecules.
The definition "catalytic fibrils" has emerged very recently, referred to amyloid fibrils on which an enzyme is immobilised [145], although catalytically active fibrils were not produced so far.
R.B. Wickner and coworkers, studying Ure2p inactivation upon aggregation, found that, when GST was fused to the fibrillogenic domain of Ure2p, a decreased enzymatic activity was detected upon fibrils formation. Similar results were obtained when the fibrillogenic domain was fused to enzymes such as barnase, carbonic anhydrase or green fluorescent protein (GFP) [156]. Finally, T.R. Serio *et al.* found that GFP fused to the C-terminus of yeast Sup35 is able to self-associate in fibrillar structures harbouring green fluorescence [157].
We planned to generate new catalytic fibrils, whose monomeric unit is a fusion protein, generated by linking [1-93]ApoA-I to the GST protein. The former will act as the fibrillogenic moiety, the latter as the catalytic moiety. GST is a good candidate to generate catalytic fibrils, due to its well known use for biotechnological applications, such as biotransformations and detoxification processes.
The fusion protein was expressed and isolated as a homogeneous product; when its catalytic activity was analysed, it was found to fully retain the enzymatic activity of GST enzyme.
The fibrillogenic potential of the fusion protein was extensively investigated using a multidisciplinary approach. Upon incubation in slightly acidic conditions, the fusion protein was found to bind ANS, ThT and CR dyes, indicating that a pH-driven transition occurs that triggers the process of aggregation in amyloid-like structures.
Upon prolonged incubation under acidic conditions, the formation of insoluble material was observed. AFM analyses indicated the in this heterogeneous material typical amyloid protofilaments were present.

The fibrillar material was then trapped on suitable filters. When the GST reaction mixture was let to pass through the filter, the expected product was generated.
Taken together, our results showed that a chimeric protein, harbouring an enzymatic activity and a fibrillogenic domain, was able: (i) to form amyloid fibrils through molecular self-assembly; (ii) to retain the catalytic activity of the enzymatic moiety. Hence, amyloid structures may act as self-assembled matrices, providing a polymeric support to a catalytically active enzyme.
Fibrils may represent a new tool to generate immobilized enzymes. As amyloid fibrils are thought to be solvent filled nanotubes [180], studies will be performed to investigate their potential as active matrices, able to crowd enzyme molecules, thus enhancing the product production.

BIBLIOGRAPHY

[1] J. M. Yon (2002) J. Cell. Mol. Med. 6(3): 307-327
[2] Anfinsen CB (1973) Science, 181: 223-230
[3] Levinthal C (1968) J. Chim. Phys. 65: 44–45
[4] Weinreb et al. (1996) Biochemistry 35: 13709–13715
[5] Uversky V.N. (2002) Protein Science, 11:739-756
[6] Schweers et al. (1994). Biol. Chem. 269: 24290–24297
[7] Wright and Dyson (1999) J. Mol. Biol. 293: 321–331.
[8] Dunker et al. (2001) J. Mol. Graph. Model. 19: 26–59
[9] Uversky et al (2000) Prot Struct Func Gen 41; 415-27
[10] Schulz (1979) Molecular mechanism of biological recognition (ed., M. Balaban), pp. 79–94. Elsevier/North-Holland Biomedical Press, New York.;
[11] Whitesides G.M., Bonheva M. (2002) Proc. Natl. Acad. Sci. USA 99:4769-74
[12] Sanders A. et al. J. Mol. Biol. (2004) 336, 165–178
[13] Thornton J.M. et al. (1999) J. Mol.Biol. 293333-42
[14] Wolynes P.G., (1995) Science, 267: 1619-1620
[15] Baum et al. (1989) Biochemistry 28: 7–13
[16] Bracken (2001) J. Mol. Graph. Model. 19: 3–12.
[17] Semisotnov et al. (1991) Biopolymers 31: 119–128.
[18] Uversky et al. (1996) Biophys. Chem. 60: 79–88.[
[19] Yon J.M. (1997) Cellular and Molecular Life Sciences, 53: 557-567.
[20] Worbs et al. (2000) EMBO J. 19: 807–818
[21] Adler et al. (1973) Methods Enzymol. 27: 675–735.
[22] Fasman (1996) Circular dichroism and conformational analysis of biomolecules. Plenum Press, New York.
[23] Kelly and Price (1997) Biochim. Biophys. Acta 1338: 161–185
[24] Fujio et al. (1985) J. Biochem. (Tokyo) 98: 949–962
[25] Kayed et al. (1999) *J. Mol. Biol.* 287: 781–796
[26] Jahn T.R., Radford (2007) S.E. Arch Biochem and Biophys doi:10.1016/j.abb.2007.05.015
[27] Goldberg et al. (199) Biochem 30:2790-97
[28] Chiti, F. C.M. Dobson et al., Nature Struct. Biol. 9 (2002), pp. 137–143.
[29] J.D. Harper, P.T. Lansbury, Annu. Rev. Biochem. 66 (1997), 385–407.
[30] Jarret et al. (1993) Cell 73055-8
[31] Lomakin A et al. Proc Nat Acad Sci USA (1997) 94:7942-7
[32] Carulla N. et al. (2005) Nature 436: 554-58
[33] A.A. Serag et al. (2002) Nature Struct. Biol. 9: 734–739
[34] Y. Liu, D. Eisenberg et al. (2001) Nature Struct. Biol. 8: 211–214.
[35] Z. Guo and D. Eisenberg, (2006) Proc. Natl. Acad. Sci. 103: 8042–8047.
[36] R.A. Staniforth et al. (2001) EMBO J. 20: 4774-4781.
[37] Lomas et al (2001), Nature 357:605-07
[38] Dobson C.M. (2001) Philos. Trans R. Soc. London B. Biol. Sci. 356:133-145
[39] Butterfield DA and Kanski J. Mech. Aging. Dev. (2001) 122:945-62
[40] Forloni G. et al. (2002) Neurobiol. Aging. 23:957-76
[41] Barral J.M. et al. (2004) Semin. Cell. Dev. Biol. 15;17,29
[42] Wacker JL et al. (2004) Nat. Struct. Mol. Bio. 11:1215-22
[43] Ellis RJ, (2001) Curr. Op. Struct. Biol. 11:114-119 2001
[44] P. Westermark, M, Sipe J.D. et al, Amyloid 12 (2005), pp. 1–4

[45] J.D. Sipe and A.S. Cohen, J. Struct. Biol. 130 (2000), pp. 88–98
[46] E.D. Eanes, G.G. Glenner, J. Histochem. Cytochem. 16 (1968), pp. 673–677
[47] W.S. Gosal, S.L. et. al Protein Pept. Lett. 13 (2006), pp. 261–270.
[48] M.R. Nilsson, Methods 34 (2004), pp. 151–160
[49] F. Chiti and C.M. Dobson, Annu. Rev. Biochem. 75 (2006), pp. 333–366.
[50] M. Stefani, Biochim. Biophys. Acta 1739 (2004), pp. 5–25
[51] Carrel L.W., Lomas D.A., (1997) Lancet 350:1348
[52] M. Stefani and C.M. Dobson, J. Mol. Med. 81 (2003), pp. 678–699
[53] Perutz F.M. Nature 1997: 385:773-5
[54] Dobson C.M. Semin. Cell. Dev. Biol. 15:3-16 (2004).
[55] Kreplak L. and Aebi U., (2006) Adv Prot Chem, 73:217-33
[56] Krawczack M et al. (2000) Hum Genet, 107:362-365
[57] V Bellotti et al. (2007) Ann Med 39:200-7
[58] Johnson R.J. et al., J. Mol. Bio. (2005) 352:823-36
[59] Merlini G, Bellotti V, (2003) N. Eng. J. Med. 349:583-96
[60] Kelly J.W. (1996). Current Opinion in Structural Biology, 6: 11-17
[61] Fowler D.M. et al (2007) Trend Biotechnol 32 (5): 217-224
[62] Fernandez-Escamilla A.M. at all. Nature Biotechnology (2004): 22, 1302-06
[63] M. Bucciantini, Nature 416 (2002), 507–511
[64] Mackay J.P. et al., Structure 9 (2001) 83–91
[65] D.M. Fowler et al. PLoS Biol. 4 (2006), p. e6.
[66] M.R. Chapman et al., Science 295 (2002), 851–855
[67] D. Claessen et al., Genes Dev. 17 (2003), pp. 1714–1726.
[68] V.A. Iconomidou et al, FEBS Lett. 479 (2000), pp. 141–145
[69] P. Butko et al., Biochem. Biophys. Res. Commun. 280 (2001), pp. 212–215.
[70] Genschel J. et al FEBS Letter 430 (1998) 145-149
[71] Wotton J.C. and Federhen S. (1996) Methods Enzymol 266:554-71
[72] Gast K. et al (1995) Biochem 34:13211-218
[73] W.C. Wigley, et al. Nature Struct. Biol. 9 (2002), pp. 381–388
[74] A. Steward, S. Adhya and J. Clarke, J. Mol. Biol. 318 (2002), pp. 935–940
[75] C. Parrini, et al. Structure 13 (2005), pp. 1143–115
[76] JS Richardson, DC Richardson, Proc. Natl. Acad. Sci. USA 99 (2002) 2754-59
[77] B.M. Broome and M.H. Hecht, J. Mol. Biol. 296 (2000), pp. 961–968
[78] Chiti F. et al Proc. Natl. Acad. Sci. USA (2002) ;99 Suppl 4 16419-26
[79] Prusiner S.B., Science, 216: 136-144, (1982)
[80] J. Collinge, Annu. Rev. Neurosci. 24 (2001), pp. 519–550.
[81] K. Si, S. Lindquist and E.R. Kandel, Cell 115 (2003), pp. 879–891
[82] M. Tanaka, et al. Nature 428 (2004), pp. 323–328
[83] Satpute-Krishnan and T.R. Serio, Nature 437 (2005), pp. 262–265
[84] L. Obici et al. (2006) Amyloid 13(4):191-205
[85] Fielding J. and Fielding PE (2005) Lipid Res. 36, 211-228.
[86] Pastore L., et al. (2004) Gene 327, 153-160.
[87] Scanu A.M. et al. (1984) J. lipid. Res. 25:1593-1602
[88] Briton EA Arter Thromb. Vasc. Biol. (1996) 16;431-440
[89] Mooradian AD (2006) Metabol 55:159-167
[90] Rogers DP et al. (1998) Biochemistry 37, 11714–11725
[91] Glass C.K. et al. (1983) J. Biol. Chem. 258: 7161-67
[92] Frank P.G. and Marcel Y.L. (2000) Lipid Res. 41, 853-872.
[93] Brouillette, C. G., Anantharamaiah, G. M. (1995) Biochim. Biophys.Acta 1256,
[94] Ajees, A. A. et al. (2006) Proc. Natl. Acad. Sci. U. S. A. 103, 2126–2131

[95] Borhani, D. W., et al. (1997) Proc. Natl. Acad. Sci. U. S. A. 94, 12291–12296
[96] Sorci –Thomas M.G., Thomas MJ (2002) Trends Cardiovasc. Med.;12:121-28
[97] Wiesgraber K.H. et al (1980) J. Clin. Invest. 66:901-907
[98] Joy, T et al. (2003) Clin. Biochem. 36, 641-645.
[99] Amarzguioui, M., et al. (1998) Biochem Biophys Res Commun. 242, 534-539.
[100] Lachmann, H.J., et al. Engl. J. Med. (2002) 346, 1786-1791.
[101] Obici L., et al. (1999) Am. J. Pathol. 155, 695-702.
[102] Andreola, A., et al. (2003) J. Biol. Chem. 278, 2444-2451
[103] Relini A. et al. (2004) Biochim. Biophis. acta 1690:33-41
[104] Kayed R., et al. (2002) Science 300, 486-489.
[105] Uversky V.N. Eur. j. biochem. (2002) 269:2-12
[106] Bucciantini M., et al. (2004) J. Biol. Chem. 23, 31374-82.
[107] Barrett A.J. (1981) Methods Enzymol 80771-778
[108] Grubb A. (1984) FEBS letters 170370-374
[109] Brown and Dziegielewska (1997) Prot Sci 6(1):5-12
[110] Stubbs, M. T., et al. (1990) EMBO J. 9, 1939-1947.
[111] Martin, J. R., Zerovnik, et al. (1995) J. Mol. Biol. 246, 331-343.
[112] Tate, S., et al. (1995) Biochemistry 34, 14637-14648.
[113] Grubb A. (2000) Adv. Clin. Chem. 35;63-99
[114] Yamada M. Neuropath. (2000) 20(1):8-22
[115] Gudmundsson et al (1972) Brain 95 (2):387-404
[116] Olafsson and Grubb (2000) Amylod 7;70-79
[117] Ceru et al (2005) Med. hypotheses 64(5):955-9
[118] Janowsky et al. (2001)Nat Struct Biol 8(4):316-20
[119] Bennett et al. (1994) Prot Sci 3(9):1444-63
[120] Knaus et al. (2001) Struct Biol 8(9):770-4
[121] Staniforth, R.A., et al. (2001) EMBO J. 20, 4774–4781.
[122] Jenko S. et al. (2004) Prot Struc, Func and Bioinf 55:417-425
[123] Konno T. et al. (1999) FEBS lettes 454:122-126
[124] Konno T. (2001) Prot. Sci. 10(10):2093-101
[125] Guijarro, J.I., et al. (1998) Proc. Natl. Acad. Sci. USA 95, 4224-4228.
[126] Chiti, F., et al. (1999) Proc. Natl. Acad. Sci. USA 96, 3590-3594.
[127] Abe, K., et al. (1991) Biomed. Biochim. Acta 50, 637-641.
[128] Kondo, H et al. (1991) FEBS Lett. 278, 87-90.
[129] Arai, S. et al. (1996) Adv. Exp. Med. Biol. 389, 73-78.
[130] Morris, J.A. and Cagan, R.H. (1972) Biochim. Biophys. Acta 261, 114-122.
[131] Morris, J.A. and Cagan, R.H. (1973) J. Biol. Chem. 248, 534-539.
[132] Kohmura, M., et al. (1990) Agric. Biol. Chem. 54, 2219^2224.
[133] Somoza, J.R et al. (1993) J. Mol. Biol. 234, 390-404.
[134] Tomic, M.T.,et al. (1992) J. Biomol. NMR 2, 557-572
[135] Spadaccini R. et al. (2001) J. Mol. Biol. 305,505-14
[136] Lee S.Y. et al. (1999) Bioch 38:2340-46
[137] Tancredi T. et al. (1992) FEBS letters 310: 27-30
[138] Feynman R. Eng. Sci. (1960) 22-36
[139] Drexler KE. (1981) PNAS USA78:5275-5278
[140] Reches M, Gazit E. (2003) Science. 300(5619):625-7.
[141] Scheibel T et al. (2003) Proc Natl Acad Sci U S A. 100(8):4527-32
[142] A. S. Mostaert J .Biol. Phys. (2006) 32:393–401
[143] A. S. Mostaert, S.P. Jarvis, (2007) Nanotechnology 18:1-5
[144] Shingo Kasai Pep. Sci. 76(1), 27-33

[145] Hamada D. et al. TRENDS in Biotechnology (2004), 22(2) 92-7
[146] Zhang S. Biotechnology Advances 322 20 (2002) 321–339
[147] Zhang S. (2003) Nat. Biotech, 21(10): 1171-78
[148] Kenney J.M., Eur. J. Biochem. (2002) ;269(16):4159-63
[149] Mangione P. et al. Protein Sci. (2001) 10: 187-199
[150] Ute Slotta et al. 7(2) Pages 183 - 188 Macromol Biosci (2007)
[151] W Po Foo et al. (2006) PNAS 103(25) 9428-33
[152] M. J.Winningham, D.Y. Sogah, (1997) Macromolecules 30, 862.
[153] Daniel E. Wagner et al. (2005) PNAS 102(36) 12656–12661
[154] Ruoslathi E. Annu. Rev. Cell. Dev. Biol. (1996) 12:697
[155] Schneider J.P. et al. (2002) J Am Chem Soc 2002,124:15030-37
[156] Baxa U. et al. PNAS (2002) 99, 8: 5253–60
[157] T.R. Serio et al. Methods Enzymol. 309 (1999), pp. 649–673
[158] Sambashivan S. et al. Nature. (2005) Sep 8;437(7056):197-8
[159] J.F. Smith et al., Proc. Natl. Acad. Sci. U. S. A. (2006), 103: 15806–11
[160] GST Gene Fusion System Handbook, Amersham Pharmacia GE (2004)
[161] Sambrook J. et al. (1989), Ed. Cold Spring Harbour Labratory press, N.Y.
[162] Laemmli U. Nature (London) Env Sci Thechnol (1970) 24:1086-89
[163] Sreerama, N., and Woody, R. W. (2000). Anal. Biochem. 287, 252–260
[164] Becktel W.J., Schellman J.A. (1987) Biopolymers 26:1859–1877.
[165] Robertson, A. D., and Murphy, K. P. (1997) Chem. Rev. 97, 1251–1267
[166] Myers et al., (1995) Protein Sci. 4, 2138–2148.
[167] Guruprasad, K., et al. (1990) Protein Eng. 4,155-161
[168] Chou P.Y., Fasman G.D. (1978) Adv. Enzym. 47:45-148
[169] Gasteiger E et al in John M. Walker (ed): The Proteomics Protocols Handbook, Humana Press (2005) pp. 571-607
[170] Linding R, et al Structure. (2003) Nov;11(11):1453-9
[171] Sebastian Maurer-Stroh et al, Switch, VIB Vrije Universiteit Brussel
[172] DeLano, W.L. The PyMOL Molecular Graphics System (2002) DeLano Scientific, San Carlos, CA, USA
[173] Thompson J.D. et al (1997) Nucleic Acids Res. 24, 4876-82
[174] Di Gaetano S. et al. (2006) Biochem. Biophys. Res. Commun. 351, 223-228.
[175] Arai M., Kuwajima K (2000) Adv. Prot. Chem. 53:209-82
[176] Klunk, W. E. et al. (1989) J. Histochem. Cytochem. 37, 1273–1281
[177] J Ma et. al (1994) Nature 372, 92.
[178] Levine H. (1993) Prot. Sci. , 404
[179] Naiki H et al., (1989) Anal. Biochem. 177,244
[180] Perutz MF et al .(2002) PNAS 99(8):5591-95
[181] Nagata K et al. (2000) Biochem. 39:14753-60
[182] Kaplan et al. (1997) Protein Sci. 6: (2)399–406.
[183] Armstrong R.N. (1991) Chem. Res. Toxicol. 4:131-40
[184] Teal F.W.J. (1960) Biochem. J 76:381-88
[185] Masino L. et al. (2002) FEBS Lett. 513: (2–3)267–272
[186] Abeliovic H., Shlomai J. (1995) Anal. Biochem. 228:351-54
[187] Dong J., et al. Amyloid. (2006) 13(4):206-15

List of publications inherent to the Doctorate research activity of Dr. Fulvio Guglielmi

Publications:

- Di Gaetano S., Guglielmi F., Arciello A., Mangione P., Monti M., Pagnozzi D., Raimondi S., Giorgetti S., Orrù S., Canale C., Pucci P., Dobson C.M., Bellotti V. and Piccoli R. (2006) Recombinant amyloidogenic domain of ApoA-I: analysis of its fibrillogenic potential. Biochem. Biophys. Res. Commun. 351, 223-228.

Comunications

- S. Di Gaetano, F. Guglielmi, A. Arciello, P. Mangione, M. Monti, D. Pagnozzi, S, Raimondi, S. Giorgetti, S. Orrù, D.M. Monti, A. Cozzolino, C. Canale, P. Pucci, V. Bellotti and R. Piccoli. Expression and characterization of recombinant apolipoprotein A-I (ApoA-I) fibrillogenic domain: first steps toward a structural and functional definition of ApoA-I associated amyloidosis. 51° Congresso Nazionale Società Italiana Biochimica, It. J. Biochem. abstr. 15.28, vol 55, September 2006, Riccione, Italy.
- A. Arciello, D.M. Monti, S. Di Gaetano, F. Guglielmi, L. De Rosa, A. Cozzolino, S. Caserta, A. Relini, S. Guido and R. Piccoli. La citotossicità del dominio N-terminale della apolipoproteina A-I (ApoA-I), responsabile della formazione di fibrille amiloidi. Giornate Scientifiche del Polo delle Scienze e delle Tecnologie per la Vita, Settembre 2007, Napoli, Italy.
- M. Minutolo, A. Massa, P. Chiaiese, A. Arciello, F. Guglielmi, R. Piccoli and E. Filippone. Espressione in Tabacco (Nicotiana Tabacum L.) dei geni umani APOA-I e LCAT coinvolti nel catabolismo del colesterolo. Giornate Scientifiche del Polo delle Scienze e delle Tecnologie per la Vita, Settembre 2007, Napoli, Italy.
- A. Arciello, D.M. Monti, S. Di Gaetano, F. Guglielmi, A. Cozzolino, S. Caserta, L. Obici, A. Relini, V. Bellotti, S. Guido and R. Piccoli. The 93-residue N-terminal domain of apolipoprotein A-I (ApoA-I), responsible for amyloid formation, is a cytotoxic agent. 52° Congresso Nazionale Società Italiana Biochimica, It. J. Biochem. abstr. 15.43, vol. 56, September 2007, Riccione, Italy.

Research activity in Scientific Institutions abroad

- From June to September 2007, the research activity of Dr. Guglielmi has been carried out at the Medical Research Council (MRC), National Institute For Medical Research (NIMR), London, UK.

ACKNOWLEDGEMENTS

I thank Prof. Renata Piccoli, my tutor, for her valuable support, precious advice and constant enthusiasm. In these years I learned a lot from her and she was really precious for me in every moment.

I am really grateful to Dr Annalisa Pastore for the opportunity she gave me to spend three months at the NIMR. I will never forget in my life this great experience!

I thank Prof. Pierandrea Temussi too, who gave me the possibility to work on a very interesting project.

I thank the Coordinator of the Doctorate and all Professors that worked to make unforgettable these years.

I thank Dr. Sonia Di Gaetano, Daria Monti and Angela Arciello, they were not only good colleagues, but first of all precious friends. It has been great to work with them!

I thank Elio, Pasquale, Gennaro, Marianna and Fulvia for their kindness and support during these years.

I thank all my friends of the Moleclar Structure Division as their friendship let me feel at home during my stage in the London: thank you Laura, Steve, Veronica, Giovanna, Liljana, Gianfelice, Valeria, Cesira and Beppe.

I thank my parents as they believe in me.

Finally, but not at the end of the list, I thank Dr Emilia De Lisa, to whom I dedicate this work. She always supported and helped me. She is very important for me and I think she is special!

Available online at www.sciencedirect.com
ScienceDirect
Biochemical and Biophysical Research Communications 351 (2006) 223–228
BBRC
www.elsevier.com/locate/ybbrc

Recombinant amyloidogenic domain of ApoA-I: Analysis of its fibrillogenic potential ☆

Sonia Di Gaetano [a], Fulvio Guglielmi [b], Angela Arciello [b], Palma Mangione [c], Maria Monti [d,e], Daniela Pagnozzi [d], Sara Raimondi [c], Sofia Giorgetti [c,f], Stefania Orrù [d], Claudio Canale [g], Piero Pucci [d,e], Christopher M. Dobson [h], Vittorio Bellotti [c,f], Renata Piccoli [b,*]

[a] *Istituto di Biostrutture e Bioimmagini, CNR, Napoli 80134, Italy*
[b] *Dipartimento di Biologia Strutturale e Funzionale, Università di Napoli Federico II, Complesso Universitario di Monte S. Angelo, via Cinthia 4, Napoli 80126, Italy*
[c] *Dipartimento di Biochimica, Università di Pavia, Pavia 27100, Italy*
[d] *CEINGE Biotecnologie Avanzate, Napoli 80131, Italy*
[e] *Dipartimento di Chimica Organica e Biochimica, Università di Napoli Federico II, Napoli 80126, Italy*
[f] *Laboratorio di Biotecnologie IRCCS, Pavia 27100, Italy*
[g] *Dipartimento di Fisica, Università di Genova, Genova 16146, Italy*
[h] *Department of Chemistry, University of Cambridge, Cambridge CB2 IEW, UK*

Received 3 October 2006
Available online 23 October 2006

Abstract

A variety of amyloid diseases are associated with fibrillar aggregates from N-terminal fragments of ApoA-I generated through a largely unexplored multi-step process. The understanding of the molecular mechanism is impaired by the lack of suitable amounts of the fibrillogenic polypeptides that could not be produced by recombinant methods so far. We report the production and the conformational analysis of recombinant ApoA-I 1–93 fragment. Similarly to the polypeptide isolated *ex vivo*, a pH switch from 7 to 4 induces a fast and reversible conformational transition to a helical state and leads to the identification of a key intermediate in the fibrillogenesis process. Limited proteolysis experiments suggested that the C-terminal region is involved in helix formation. The recombinant polypeptide generates fibrils at pH 4 on a time scale comparable with that of the native fragment. These findings open the way to studies on structural, thermodynamic, and kinetic aspects of ApoA-I fibrillogenesis.

Keywords: Fibrillogenesis; ApoA-I amyloidosis; Recombinant amyloidogenic proteins; Conformational analysis

Abnormal conformations of specific proteins and polypeptides represent the underlying pathogenic basis of amyloid diseases [1]. A variety of amyloid diseases are associated with mutations in apolipoprotein A-I (ApoA-I), for which the process of fibril formation has not yet been clarified, and a paucity of information exists in comparison to the detailed knowledge surrounding other disease-associated proteins such as lysozyme, transthyretin, and β2-microglobulin [1]. Analysis of natural amyloid fibrils has shown that fibrils consist of ApoA-I N-terminal fragments, 90–100 residues long. Mutations are sometimes present within the N-terminal portion of the protein that is

☆ *Abbreviations:* AFM, atomic force microscopy; ApoA-I, apolipoprotein A-I; [1–93]ApoA-I, the 93-residues N-terminal domain of ApoA-I; ANS, 8-anilino-1-naphthalenesulphonate; DTT, dithiothreitol; TEM, transmission electron microscopy; ESMS, electrospray mass spectrometry; IPTG, isopropyl-β-D-thiogalactopyranoside; MALDI-MS, matrix assisted laser desorption ionization mass spectrometry; TFE, trifluoroethanol.

* Corresponding author. Fax: +39 081 679233.
E-mail address: piccoli@unina.it (R. Piccoli).

0006-291X/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.bbrc.2006.10.026

eventually found in fibrils (''internal mutations''), but can also occur in positions located outside this region of the polypeptide sequence (''external mutations'') [2].

Characterization of the polypeptides purified from natural fibrils from patients carrying the ''external mutation'' Leu174Ser has shown that the 1–93 N-terminal portion of ApoA-I is intrinsically amyloidogenic in a physiological environment. By using the natural 1–93 polypeptide, we have shown that acidic conditions (pH 4) induce fibril formation [3]. We envisage a complex aggregation pathway in which the polypeptide assumes a random coil structure at neutral pH, shifts into an unstable helical conformation at acidic pH, and then aggregates into a β-sheet-based polymeric structure. The investigation of this complex pathway has been limited so far by the small amounts of the natural amyloidogenic polypeptide. Attempts to express an homologous recombinant form either in prokaryotic or eukaryotic cells have failed, probably for the rapid intracellular digestion of the unstructured product.

We describe here the production of recombinant 1–93 polypeptide and the investigation of its structural dynamics during fibrillogenesis.

Materials and methods

Expression and isolation of Apo A-I 1–93 polypeptide. The cDNA encoding fragment 1–93 of ApoA-I ([1–93]ApoA-I) was obtained by PCR amplification using full length ApoA-I cDNA as a template, and oligonucleotides 5′-CGCGGATCCGATGAACCCCCCCAGAG-3′ (forward primer) and 5′-CCGGAATTCTTACACCTCCTCCAGATCCTTG-3′ (reverse primer), where restriction sites are underlined. The amplified DNA was cloned in the *Eco*RI and *Bam*HI sites of pGEX-4T-3 expression vector (General Electric, CT) downstream to the sequence encoding glutathione *S*-transferase (GST). Competent *Escherichia coli* BL21DE3 cells were transformed and induced to express the recombinant protein by the addition of 0.1 mM IPTG for 2 h at 37 °C, using a BioFlo 3000 benchtop fermentor (New Brunswick Scientific, NY).

Isolation of recombinant [1–93]ApoA-I. Lysates were obtained upon treatment of bacteria with 1% Triton X-100 in PBS (30 min, 4 °C) in the presence of protease inhibitors (Roche, Germany), followed by sonication (Misonix, Farmingdale, NY). The soluble fraction obtained by centrifugation was fractionated by affinity chromatography on a GSTrap glutathione-agarose column. GST-containing species were eluted with 50 mM Tris–HCl at pH 8.0, containing 10 mM glutathione following the manufacturer's protocol, but in the absence of DTT to avoid inhibition of thrombin enzymatic activity. Proteins were then digested with 6 U/mg of thrombin (Sigma) for 30 h at 4 °C in the chromatography elution buffer. Products were separated by HPLC reverse chromatography on a Ultrapure C_8 column (Beckman Coulter, CA) with a linear gradient of acetonitrile in 5 mM phosphate buffer at pH 7.4 and analysed by SDS–PAGE on 15% acrylamide gels. For Western blot analyses anti-human ApoA-I polyclonal antibodies (DAKO, Denmark) and a chemiluminescence detection system (West Pico, Pierce) were used. Pure [1–93]ApoA-I was dialyzed in water, lyophilized, and stored at −70 °C until use. The lyophilized polypeptide was dissolved in phosphate buffer at pH 7.4. About 1 mg of pure [1–93]ApoA-I was obtained from 1 L of bacterial culture.

In situ hydrolysis and MALDI-MS analysis. The protein band stained by Coomassie blue brilliant and corresponding to the expected recombinant protein was excised and *in situ* digested according to [4] and the peptide mixture was analysed by MALDI as described [5].

Limited proteolysis. Aliquots of 7 μM [1–93]ApoA-I were digested separately with trypsin, V8 protease, chymotrypsin, elastase, and subtilisin. Proteolysis was performed at 25 °C in 12 mM sodium phosphate at pH 7.5, using enzyme-to-substrate ratios ranging between 1:800 and 1:4000 (w/w). Proteolysis in the presence of 20% TFE was performed under the same conditions using enzyme-to-substrate ratios ranging from 1:100 to 1:2000 (w/w). The extent of proteolysis was monitored by sampling the incubation mixture at different time intervals. Proteolytic fragments were fractionated and characterized by mass spectrometry as described [5].

Gel-filtration assay. Gel-filtration experiments were performed with a SMART system (Pharmacia Biotech) using a Superdex-75 column (Pharmacia) equilibrated and eluted in 50 mM phosphate buffer at pH 7.5, containing 150 mM NaCl, in the presence or absence of 20% TFE. [1–93]ApoA-I (4 μg) was dissolved in an appropriate volume of buffer with or without TFE (20%) to a final concentration of 7 μM. Experiments were performed at room temperature at a flow rate of 75 μl/min. The column was calibrated under the same conditions (either with or without 20% TFE) with standard proteins.

CD spectra. Circular dichroism spectra were recorded on a Jasco J-710 spectropolarimeter as previously described [3]. Measurements were performed at 20 °C at a protein concentration of 0.3 mg/ml in 3 mM glycine, 3 mM sodium acetate, and 3 mM sodium phosphate at pH 7 (buffer A). Acidification and neutralization of the solution were carried out as previously described [3]. CD data were expressed as mean residue ellipticity (θ).

Fluorescence spectra at the equilibrium. By observing the intrinsic fluorescence emission of tryptophan in the range 300–550 nm and excitation at 295 nm, we monitored the conformational transition induced by the pH jump from 7 to 4. Measurements were carried out at 20 °C in a 10-mm cell by using a Perkin-Elmer LS50 spectrofluorimeter and the slit widths set at 5 nm. Protein solution (0.06 mg/ml) was first analysed in buffer A (see above). The solution was then analysed after acidification to pH 4 followed by neutralization to pH 7 by adding HCl and NaOH, respectively, to give a concentration of 3.8 mM in each sample.

Stopped-flow fluorescence spectra. The kinetics of the [1–93]ApoA-I structural transition from pH 7 to pH 4 were monitored at 20 °C on a Bio-Logic SFM-300 stopped flow fluorimeter using an excitation wavelength of 295 nm and monitoring the total fluorescence emission change over 320 nm. One volume of 0.6 mg/ml peptide solution in buffer A (see above) was mixed with five volumes of 5.4 mM glycine, 5.4 mM sodium acetate, and 5.4 mM sodium phosphate at pH 7, and four volumes of 0.01 N HCl.

Binding to 8-anilino-1-naphthalenesulphonate (ANS). ANS binding experiments were carried out on solutions containing [1–93]ApoA-I (2.8 μM) and ANS (330 μM) in buffer A (see above). 8-Anilino-1-naphthalenesulphonate emission fluorescence spectra were recorded in the range 400–600 nm at an excitation wavelength of 395 nm and the slit widths set at 5 nm [6]. Spectra were then recorded after acidification to pH 4, and then after returning to pH 7.

Fibrils imaging. Samples for TEM were prepared by floating the aggregate suspension on formvar/carbon-coated grids for 2–3 min, before air-drying and staining them with 2% uranyl acetate. Samples were examined in a Jeol JEM 1200 EX electron microscope operating at 180 kV. For AFM imaging, [1–93]ApoA-I solution (0.25 mg/ml) was incubated at pH 4 and 25 °C for two weeks and analysed as previously described [7].

Results

Expression and isolation of recombinant [1–93]ApoA-I as a stable product

The 1–93 fragment of ApoA-I was expressed in bacterial cells following an experimental strategy aimed at reducing the intracellular degradation of the polypeptide during its production. It was expressed as a chimeric protein obtained by fusing the 93 residues polypeptide to glutathione

S-transferase (GST). Analyses by SDS–PAGE of bacterial lysates showed the presence of a major protein species in IPTG-induced cells (Fig. 1A, lane 3) with a molecular mass of about 36 kDa, as expected for the chimeric product, and specifically recognised by an anti-ApoA-I antibody (Fig. 1B, lane 3). In this sample, additional immunopositive species with a molecular mass lower than that of the chimeric protein were also present (see below).

GST-containing proteins, selected on a GSH-agarose affinity column, were analysed by SDS–PAGE (Fig. 1A, lane 4) and found to contain exclusively species recognized by the anti-ApoA-I antibody (Fig. 1B, lane 4). Coomassie-stained protein bands corresponding to these immunopositive species were excised from the gel and digested *in situ* with trypsin. The resulting peptide mixtures were directly analysed by MALDI mass spectrometry. Mass signals were mapped onto the anticipated ApoA-I sequence and showed that the main product was the GST-[1–93]ApoA-I full length chimeric protein (indicated in Fig. 1 by an arrow). In addition to this product, small quantities of C-terminal truncated [1–93]ApoA-I species were also identified.

The [1–93]ApoA-I moiety was released from the chimeric protein by targeted proteolysis, making use of a unique cleavage site for thrombin positioned between the GST and the [1–93]ApoA-I coding sequences. The proteolysis mixture, analysed by SDS–PAGE and Western blot (Fig. 1A and B, lane 5), was resolved by reverse-phase HPLC. The major eluted fraction was found to contain a pure protein species with the same electrophoretic mobility as the native fibrillogenic polypeptide [8] recognized by the anti-ApoA-I antibody (Fig. 1A and B, lane 6). An aliquot of this protein was analysed by electrospray mass spectrometry (ESMS) showing a molecular mass of 10863.8 ± 0.1 Da, perfectly matching the expected mass value of [1–93]ApoA-I with the extra dipeptide Gly-Ser at the N-terminus originating from the plasmid construct following thrombin cleavage (10864.0 Da).

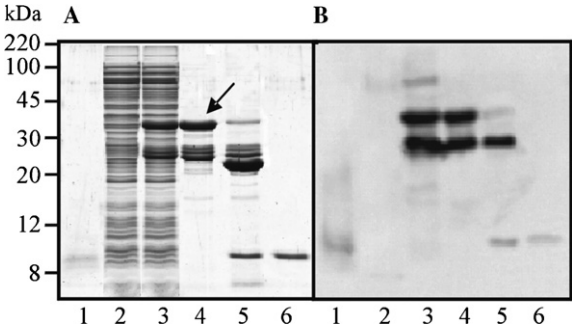


Fig. 1. Analysis by SDS–PAGE of recombinant [1–93]ApoA-I. (A) Coomassie staining; (B) Western blotting. Lane 1, native [1–93]ApoA-I extracted from *ex vivo* fibrils; lanes 2 and 3, soluble fractions prepared from non-induced cells (lane 2) or induced cells (lane 3); lane 4, GST-containing proteins selected by affinity chromatography; lane 5, thrombin proteolytic products; lane 6, HPLC purified [1–93]ApoA-I. The arrow indicates the full length GST-fused [1–93]ApoA-I.

CD spectroscopic analysis

The overall far UV CD spectrum of [1–93]ApoA-I at neutral pH (Fig. 2A), with a minimum at 203 nm, indicates that the peptide is highly unstructured under these conditions. A marked change in the spectrum is observed at pH 4, with the major minimum in the spectrum shifting from 203 to 208 nm and a considerable increase in ellipticity at 222 and 190 nm. Such spectral changes are consistent with a transition to helical structure at pH 4, and are in good agreement with the behaviour of the natural [1–93]ApoA-I polypeptide [3]. In a similar way, we have previously described the decay of the CD signal associated with protein aggregation and precipitation as well as the progressive loss of CD signal and the further aggregation when the protein is left at pH 4 [3]. The pH-induced changes in the spectrum can be reversed if NaOH is added within a few seconds to neutralize the pH (Fig. 2A). Fig. 2B reports the time dependence of CD spectra changes of [1–93]ApoA-I upon exposure to acidic conditions. After 60 min of incubation at pH 4 the spectra show the disappearance of the minimum at 222 and the persistence of a second minimum at 205 nm that is suggestive of a transition to β-sheet structure.

Fluorescence spectroscopy

Tryptophan emission fluorescence was monitored for [1–93]ApoA-I at pH 7 and 4 (Fig. 2C). A blue shift from 347 to 342 nm occurs during the pH transition and is associated with a decrease in the fluorescence intensity in complete agreement with the data obtained for the natural peptide [3]. The effect of the pH jump on the intrinsic fluorescence spectrum is consistent with the collapse of the protein core region and the exclusion of water molecules in the vicinity of aromatic residues. This transition again is fully reversible when the pH is rapidly returned to 7. The rate of the conformational transition induced by low pH was then investigated by monitoring the change in the tryptophan intrinsic fluorescence under stopped flow conditions (inset of Fig. 2C). The 5% decrease of intrinsic fluorescence emission, expected on the basis of studies carried at the equilibrium, is complete in approximately 2 s ($k = 1.01\ s^{-1}$) and 90% of this change takes place within the lag time of the data acquisition (<2 ms).

ANS binding

The binding of the apolar dye ANS to [1–93]ApoA-I is associated with an enhanced fluorescence intensity and a blue shift in the emission wavelength. Such changes are frequently used to detect partially folded intermediate states of globular proteins and are characteristic of solvent-exposed hydrophobic clusters [9]. In the case of [1–93]ApoA-I, binding of ANS to the protein at pH 7 leads to a slight blue shift from 515 to 489 nm but no increase in fluorescence emission (compare Fig. 2D with inset).

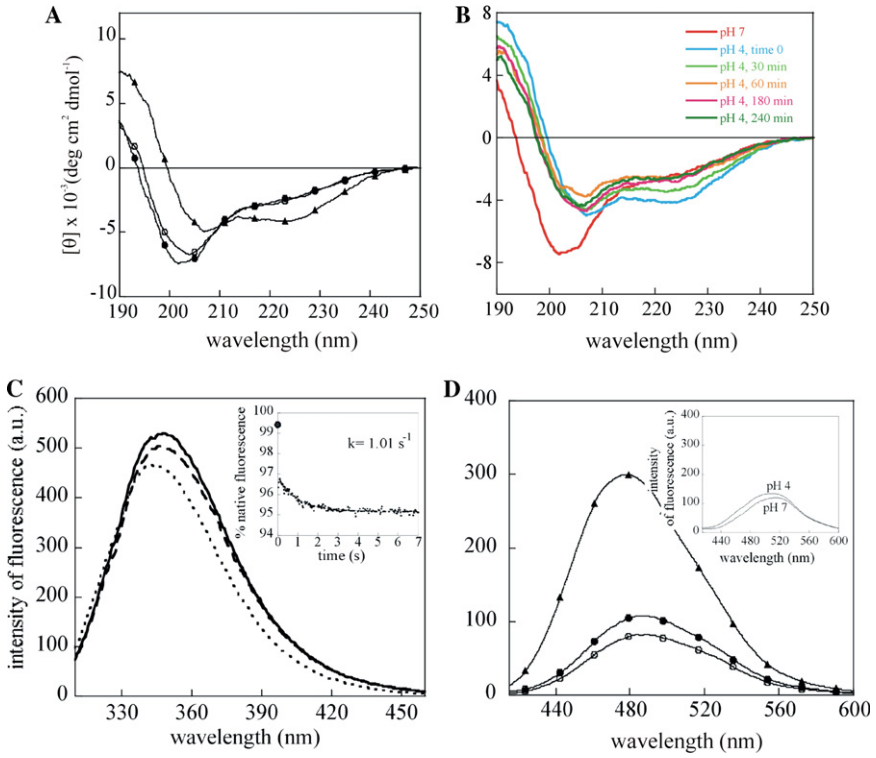


Fig. 2. Spectroscopic analyses of [1–93]ApoA-I. (A) pH-induced transition of [1–93]ApoA-I secondary structure monitored by far UV CD. Spectra were recorded at pH 7 (●), pH 4 (▲), and then returned to pH 7 (○). (B) Time dependence of the induction of secondary structure of [1–93]ApoA-I at pH 4. The far UV CD spectra were recorded at pH 7 and after a pH jump to pH 4 at the indicated time intervals. (C) Effect of pH on the intrinsic fluorescence of [1–93]ApoA-I. Tryptophan emission was monitored at pH 7 (solid line), after a pH jump to 4 (dotted line), and then a return to pH 7 (dashed line). The inset shows the rate of change of intrinsic fluorescence acquired during the pH jump. The data were normalized by defining the fluorescence of the protein at pH 7 as 100%. The continuous line through the data points represents the best fit to a single exponential function. The symbol on the abscissa refers to the fluorescence intensity recorded at $t = 0$. (D) pH-induced binding of ANS to [1–93]ApoA-I. Symbols are as (A). ANS emission fluorescence spectra were recorded in the range of 400–600 nm at the excitation wavelength of 395 nm with the slit widths set at 5 nm. Spectra of the free dye at pH 7 and 4 are shown in the inset.

Reduction in the pH, however, leads to a substantial change of ANS fluorescence reflecting the pH-induced transformation into a partially folded conformation; at pH 4 the ANS maximum emission shifts to 477 nm with a considerable increase in fluorescence intensity (Fig. 2D) consistent with the behaviour previously ascribed to proteins in a molten globule state [9].

The ANS–[1–93]ApoA-I complex at pH 4 dissociates when the pH is returned to pH 7 and this reversibility is consistent with data derived from the CD and intrinsic fluorescence spectra. The pH dependence of the fluorescence intensity of the free dye is shown for comparison in the inset of Fig. 2D.

Limited proteolysis

The conformational transitions of [1–93]ApoA-I were investigated by limited proteolysis coupled with mass spectrometry both in the presence and in the absence of TFE (20% at pH 7.5), which induces a stable helical state. The patterns of preferential proteolytic sites were obtained using a set of five proteases (trypsin, V8 protease, chymotrypsin, elastase, and subtilisin) as conformational probes [5]. For each protease, the appropriate enzyme/protein ratio was carefully determined to generate a limited number of proteolytic events directed to the most flexible and solvent-exposed regions of the polypeptide chain. Proteolytic fragments were separated by reverse-phase HPLC and identified by ESMS, allowing the positions of the cleavage sites to be assigned.

Fig. 3A summarizes the overall data from limited proteolysis experiments. In the absence of TFE, the polypeptide chain is preferentially cleaved in its central region, between residues 17 and 61, whereas the N-terminal region (residues 1–16) and the large segment 62–93 at the C-terminus are protected. Similar results were obtained when the proteolysis experiments were carried out in the presence of 20% TFE.

Gel filtration

The oligomeric state of the amyloidogenic [1–93]ApoA-I fragment was assessed by gel-filtration chromatography at pH 7.5 both in the presence and in the absence of 20% TFE. Interestingly, under native conditions the elution volume of [1–93]ApoA-I species was consistent with a dimeric

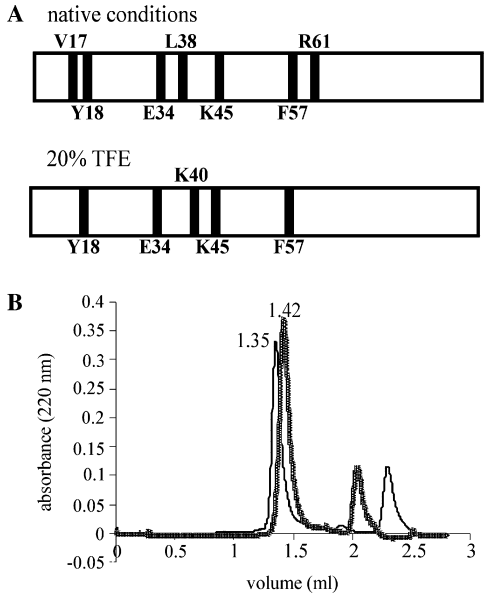


Fig. 3. (A) Schematic representation of the results obtained by limited proteolysis experiments. Preferential proteolytic sites occurring in [1–93]ApoA-I in the absence (native conditions) and presence of 20% TFE are indicated by solid bars. (B) Investigation of the oligomeric state of [1–93]ApoA-I by gel-filtration chromatography, in the absence (thin line) or presence (heavy line) of 20% TFE. The corresponding elution volumes are indicated.

structure, whereas in the presence of the co-solvent the protein showed the behaviour of a monomeric species (Fig. 3B).

Fibril formation

The formation of amyloid fibrils at pH 4 was monitored by electron microscopy and atomic force microscopy. Fig. 4A shows an electron microscopy image of a fibrils nest obtained with recombinant [1–93]ApoA-I. Images obtained by AFM are presented in Fig. 4B. Narrow protofilaments and well-defined fibrils were detectable after 72–90 h of incubation at pH 4, consistent with the rate observed for the native protein [3]. Fibrillar structures coexist with globular aggregates whose height is between 4 and 10 nm. The heights of protofilaments and mature fibrils are 0.8 ± 0.3 nm and 2.4 ± 0.6, values in agreement with those measured by tapping mode AFM in air for *ex vivo* amyloid fibrils formed by Leu174Ser ApoA-I [7]. The relatively small height values measured by AFM in air are associated with sample drying, which results in dehydration and possibly deformation of the fibrils upon adhesion to the mica substrate. On the other hand, for Leu174Ser ApoA-I fibrils, a diameter of about 7 nm was estimated by electron microscopy [10] and a height value of 11 nm was measured by AFM under fully hydrated conditions [7].

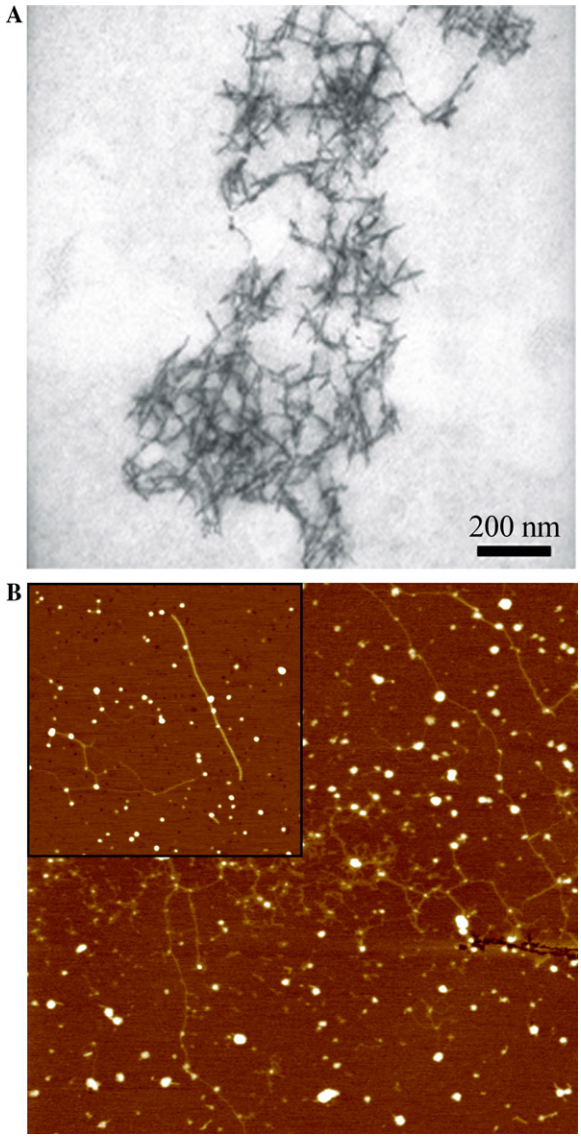


Fig. 4. Microscopic analyses of aggregated [1–93]ApoA-I. (A) Transmisson electron microscopy image of [1–93]ApoA-I fibrils stained by 2% uranyl acetate. (B) Tapping mode atomic force microscopy image of [1–93]ApoA-I fibrillar material showing a network of thin filaments coexisting with globular aggregates. Scan size, 4.0 μm; *Z* range, 15 nm. Inset, thin filaments and a fibril approximately 1 μm long. Globular aggregates are also present. Scan size, 1.8 μm; *Z* range, 10 nm.

Discussion

In the present study an efficient strategy to produce a recombinant version of the 1–93 amyloidogenic domain of ApoA-I has been established. The investigation of the conformation of this polypeptide in solution indicates that in physiological conditions the protein has a highly unfolded structure. A decrease in the pH value from 7 to 4 induces a predominant α-helical structure, by inducing the conversion of the protein from a random coil to a helical/molten globule state, capable to bind ANS. This transition, complete within 2 s and fully reversible when the pH is returned to 7, is followed by the appearance of a significant β-sheet component. The helical/molten globule intermediate displays a strong propensity to oligomerize; indeed attempts to analyse this intermediate by gel-filtration failed due to the rapid formation of insoluble species (data not shown).

The conformation of [1–93]ApoA-I polypeptide in the helical state was further investigated by stimulating the helical conformation at neutral pH by treatment with 20% trifluoroethanol [11]. Limited proteolysis experiments, performed both in the absence and in the presence of the co-solvent, indicated a nearly identical distribution of preferential proteolytic sites, with the C-terminal region of the polypeptide chain being inaccessible to proteases in both conditions. Gel-filtration suggests that [1–93]ApoA-I, while dimeric in native conditions, is monomeric in the presence of TFE. The inaccessibility of the C-terminus in native conditions might then be a consequence of the dimeric structure of the polypeptide, whereas the local stabilization induced by TFE makes this region resistant to proteolytic attack in the monomeric form.

Analyses by electron and atomic force microscopy have demonstrated that [1–93]ApoA-I generates typical amyloid fibrils following incubation at pH 4 for lengths of time comparable to those required by the natural counterpart.

Our results on recombinant [1–93]ApoA-I confirm and expand the data obtained with its natural counterpart and focus our attention on the kinetics and stability of the helical/molten globule state, as a key intermediate in the multistep fibrillogenic process. These findings also encourage the production of the variants of this polypeptide described in patients affected by ApoA-I associated amiloidoses. A detailed comparative study on the conformational dynamics and fibrillogenic potential of mutated forms versus wild type [1–93]ApoA-I would contribute greatly to the comprehension of the molecular mechanism by which mutations in ApoA-I promote amyloid deposition.

Acknowledgments

This study was supported by MIUR, Ministero dell'Università e della Ricerca Scientifica (PRIN 2005053998, FIRB 2001 project RBNE01S29H, and FIRB 2001 project RBAU015B47). Ricerca finalizzata Ministero della Salute, Fondazione Cariplo (Progetto Nobel), and by Regione Campania (2002). The research of C.M.D. is supported by Programme Grants from the Wellcome and Leverhulme Trusts.

References

[1] F. Chiti, C.M. Dobson, Protein misfolding, functional amyloid, and human disease, Annu. Rev. Biochem. 75 (2006) 333–366.
[2] L. Obici, G. Franceschini, L. Calabresi, S. Giorgetti, M. Stoppini, G. Merlini, V. Bellotti, Structure, function and amyloidogenic propensity of apolipoprotein A-I, Amyloid 13 (2006) 1–15.
[3] A. Andreola, V. Bellotti, S. Giorgetti, P. Mangione, L. Obici, M. Stoppini, J. Torres, E. Monzani, G. Merlini, M. Sunde, Conformational switching and fibrillogenesis in the amyloidogenic fragment of apolipoprotein A-I, J. Biol. Chem. 278 (2003) 2444–2451.
[4] A. Shevchenko, P. Keller, P. Scheiffele, M. Mann, K. Simons, Identification of components of *trans*-Golgi network-derived transport vesicles and detergent-insoluble complexes by nanoelectrospray tandem mass spectrometry, Electrophoresis 18 (1997) 2591–2600.
[5] A. Scaloni, N. Miraglia, S. Orru, P. Amodeo, A. Motta, G. Marino, P. Pucci, Topology of the calmodulin-melittin complex, J. Mol. Biol. 277 (1998) 945–958.
[6] D.P. Rogers, C.G. Brouillette, J.A. Engler, S.W. Tendian, L. Roberts, V.K. Mishra, G.M. Anantharamaiah, S. Lund-Katz, M.C. Phillips, M.J. Ray, Truncation of the amino terminus of human apolipoprotein A-I substantially alters only the lipid-free conformation, Biochemistry 36 (1997) 288–300.
[7] A. Relini, R. Rolandi, M. Bolognesi, A. Gliozzi, M. Aboudan, G. Merlini, V. Bellotti, Ultrastructural organization of ex-vivo amyloid fibrils formed by the apolipoprotein A-I Leu174Ser variant: an atomic force microscopy study, Biochim. Biophys. Acta 1690 (2004) 33–41.
[8] L. Obici, V. Bellotti, P. Mangione, M. Stoppini, E. Arbustini, R. Verga, I. Zorzoli, E. Anesi, G. Zanotti, C. Campana, M. Viganò, G. Merlini, The new apolipoprotein A-I variant $Leu^{174} \rightarrow Ser$ causes chardiac amyloidosis, and the fibrils are constituted by the 93-residue N-terminal polypeptide, Am. J. Pathol. 155 (1999) 695–702.
[9] G.V. Semisotnov, N.A. Rodinova, O.I. Razgulyaev, V.N. Uversky, A.F. Gripas, R.I. Gilmanshin, Study of the "molten globule" intermediate state in protein folding by a hydrophobic fluorescent probe, Biopolymers 31 (1991) 119–128.
[10] P. Mangione, M. Sunde, S. Giorgetti, M. Stoppini, G. Esposito, L. Gianelli, L. Obici, L. Asti, A. Andreola, P. Viglino, G. Merlini, V. Bellotti, Amyloid fibrils derived from the apolipoprotein A-I Leu174Ser variant contain elements of ordered helical structure, Protein Sci. 10 (2001) 187–199.
[11] M. Buck, Trifluoroethanol and colleagues: cosolvents come of age. Recent studies with peptides and proteins, Q. Rev. Biophys. 31 (1998) 297–355.